Robots that Say 'No': Acquisition of Linguistic Behaviour in Interaction Games with Humans

Frank Förster

A thesis submitted to the University of Hertfordshire
in partial fulfilment of the requirements
of the degree of

Doctor of Philosophy

The programme of research was carried out in the School of Computer Science, Faculty of Science, Technology and Creative Arts, University of Hertfordshire.

September 2013

Dedications

To my parents Irene and Manfred, who allowed me to pursue my dreams and for their unconditional trust. To Hannah for giving me shelter and love in times when I needed it most. And to Professore Mongardi, “the last Hegelian”, who taught me true philosophy.

Acknowledgements

I would like to thank my supervisors, Chrystopher Nehaniv and Joe Saunders for their ideas, support and software. Without their help many aspects of this thesis could not have been realized.

I would also like to thank my collegues and friends from the Adaptive Systems Research Group for having created such a great research and, equally important, non-research atmosphere. Thank you Moritz, Frank, Antoine, Dag, Josh, Philipe, Joe and all the others that came and went.

Finally I would like to thank Hannah for proof-reading this creation and more importantly being so patient with me.

The ITALK project (EU Integrated Project ITALK ("Integration and Transfer of Action and Language in Robots") funded by the European Commission under contract number FP-7-214668) played a pivotal role for the making of the work described herein, far beyond the mere funding. I would especially like to thank the developer of the control software, Ugo Pattacini, for his quick help. Without this software the experiments would not have been possible. I'd like to thank the other ITALK members for some great discussions and other forms of life.

Frank Förster

Abstract

Negation is a part of language that humans engage in pretty much from the onset of speech. Negation appears at first glance to be harder to grasp than object or action labels, yet this thesis explores how this family of 'concepts' could be acquired in a meaningful way by a humanoid robot based solely on the unconstrained dialogue with a human conversation partner. The earliest forms of negation appear to be linked to the affective or motivational state of the speaker. Therefore we developed a behavioural architecture which contains a motivational system. This motivational system feeds its state simultaneously to other subsystems for the purpose of symbol-grounding but also leads to the expression of the robot's motivational state via a facial display of emotions and motivationally congruent body behaviours.

In order to achieve the grounding of negative words we will examine two different mechanisms which provide an alternative to the established grounding via ostension with or without joint attention. Two large experiments were conducted to test these two mechanisms. One of these mechanisms is so called *negative intent interpretation*, the other one is a combination of physical and linguistic prohibition. Both mechanisms have been described in the literature on early child language development but have never been used in human-robot-interaction for the purpose of symbol grounding.

As we will show, both mechanisms may operate simultaneously and we can exclude none of them as potential ontogenetic origin of negation.

Contents

In the beginning was the Word, and the Word was
with God, and the Word was God

—*John 1:1*

Chapter 1

Introduction

This thesis is about a particular subset of human speech that came to be labelled *negation*. More precisely, this thesis is about the acquisition of this particular part of human language by a humanoid robot, and by means of having this robot interact with naïve participants. With naïve participants we refer to humans that are not the constructors of the robot nor of its software architecture, and that do not know anything about the internal workings of the system. Additionally: these humans did not even know why we invited them to talk to the robot. By means of linking and showing the link between negation and emotion/motivations, the major topic of this thesis, this work connects to the more comprehensive theme of human interaction and the mutual dependence of language and emotion.

The basic idea which is elaborated upon in this thesis is roughly as follows:

1. When humans interact via communication this often consitutes a (social) action and this action cannot be reduced to or sufficiently characterised as an (act of) reference to entities outside of the interactant such as physical objects, events, other interactants, or cultural artefacts.

2. Language use, i.e. the production of words and strings of words with a certain intonational contour and energy or the production of gestures, is one of several forms of human communication.

3. Human action is motivated, i.e. humans do not (socially) act randomly but human action can be explained by refering to the actors' volition, motivation, emotion, affect, or drives - we use *motivation* as an umbrella term to refer to all of these 'things'. The relationship between action and motivation is not only explanatory but also biological: we postulate some kind of neural correlates of these phenomena, and connectivity between these correlates in terms of which it can be explained why humans do what they do, including *doings* of the linguistic kind.

4. The production of words is part of language use and therefore part of human action. This includes single negation words but also grammatically more complex negative constructions.

5. Because action is motivated we postulate biological link(s) between motivation and language including link(s) between motivation and negation.

6. (From a developmental perspective emotion as opposed to "pure cognition" is particularily prevalent and important in mother-child communication. For this reason we postulate a particularily strong link between a toddler's motivation and his or her early forms of language use, this link is mirrored in the affective valence of the first words.)

Regarding assertion 1: *Social* is bracketed because the reader may associate this word with developmentally late conventional uses of language such as speech acts to establish conventional social facts and relations (oaths, delarations of marriage, war declarations, judges'

sentences, etc.). Our use of *social* is meant to include communicative actions that are closer to biology and developmentally earlier such as a baby's cry or early communicative acts of rejection by emotional displays. We *do* take these acts, and human communication in general, to be an inherently social phenomenon. It is just by virtue of for example speech act theory having been developed 'the other way around', i.e. starting with and emphasising highly conventional and 'abstract' uses of language, that the reader might have a more narrow conception of *social* than the one we elude to here.

With 'cultural artefacts' we mean 'culturally defined objects' such as *schools*, *countries*, *TV shows*, *stories*, *nationalities*, *the military*, social roles such as *judges*, *policemen*, *protesters*, *dictators*, *the Queen*, or other somewhat 'abstract' entities that we regularly refer to in talk but which are not mainly 'defined'[1] in terms of physical properties or affordances.

Regarding assertion 2: This point is made as it is not uncommon in linguistic but also technical circles to reduce language to a linguistic code, i.e. word forms and the grammatical relations between words. We strongly oppose this attempt and emphasize the non-codified, physical, and biological characteristics of spoken language.

Regarding assertion 3: The reason for giving this list of motivation-related terms, that admittedly seems somewhat random, is the circumstance that there is not much consensus in the scientific community, which of these notions are 'biologically real', which of them are 'psychological constructs', and which are epiphenomena of our language that, on the biological level, can be reduced to one or more of the other notions. Neither is there much

[1]Is is another misconception to think that one would have to know the definition of a word in order to use it. We regularily use words without being able to give a definition of them. This author, for example, has a lamentably small knowledge of different tree species and possibly could not tell apart a spruce from a pine tree. Nevertheless the author has no problem with using (and understanding) these words in a dialogue. The problems only start if the talk ventures towards the botanical. So it would be more correct (but also bulky) to say ".. which, if one tried to define these entities one would define them ..".

agreement, if one accepts 2 or more of these notions as being 'biologically real', how these relate to each other, causally or other. Thus, the obvious vagueness of the author is indeed, at least partially, caused by the lack of agreement in psychology, neuroscience, discourse analysis, and ethnopragmatics (cf. Lindquist et al. 2013).

Regarding assertion 4: The reason why we include grammatical constructions here is our belief that the interdependence of motivation and language does not 'go away' with the progressing development of a child's language use. The interdependence might become slightly weaker with certain linguistic constructions as 'pure cognition' and grammatical skills of the child develops. But we think it to be a fatal mistake to believe that humans develop from being 'purely emotional beings' into being 'purely rational beings' and to construe adult language use as a mainly rational and possibly logical endeavour. The author perceives this 'logical stance' to be a remote echo of positivism which was partially rejuvenated with Chomsky's version of linguistics. This stance can only be upheld under ignorance of mounting empirical evidence to the contrary and goes to show that parts of science, at least in the cases of (theoretical) linguistics, does not necessarily live up to its own criteria, i.e. that theories have to be abandoned when faced with empirical evidence that contradicts the theoretical assertions.

Regarding assertion 6: This assertion is bracketed as it makes use of the dichotomy *cognition* vs. *motivation* or *cognition* vs. *emotion*. We regard this dichotomy as problematic, as it might contribute to us thinking that there were a fundamental difference between 'to think' as opposed to 'to feel'. By extending this dichotomy one might arrive at the Chomskians' believe that our mind was modular in the sense that the 'language faculty' would operate and could be studied in isolation of other 'mental' faculties. Other concoctions that can be construed based on this dichotomy are Chomsky's postulated and poorly defined divide between "I-" and "E-language", which is in strong contradiction of

empirical evidence that documents the impact of 'pragmatic' phenomena ("E-language") upon certain grammatical constructions ("I-language"). We nevertheless decided to make use of this common dichotomy as it still in heavy use in the scientific community. But we also want to emphasize the danger of its frequent use in that one might come to believe in the biological reality of this division just by virtue of having alluded to it by 'speaking it out' too often. This then would be a good example of our understanding of Wittgenstein's saying of what happens when "language celebrates"[2].

A Warning: Beware of universals Before we go into any further elaboration of the topic proper we would like to sensitise the reader to a fundamental issue that is much too often not discussed, mentioned, or made explicit in the literature on language acquisition (including robotics). This issue has to do with the clash of paradigms in current linguistics, cognitive versus nativist linguistics, and the historical impact of the older of the two paradigms, nativism, on other language-related fields such as pragmatics, developmental psycholinguistics, or developmental psychology.

The single most dangerous and tempting tendency within the sciences of language[3] is the tendency to adopt a premature assumption about what is and what is not universal across all languages. Assumptions of universality, albeit different from those in linguistics, are also common in the literature on language acquisition. To make such an assumption is

[2]This is a reference to Wittgenstein's dictum that "philosophical problems arise when language celebrates" which he attributes to the philosophers' use of language, especially in the context of ostensive definitions of words and the problems that arise if all words are ought to be given meaning via ostension including ostensive words such as "this" themselves Wittgenstein (1984) (PI 38). Anscombe, the English translator, translated the original "feiert" (*celebrates*) to "goes on holiday" in Wittgenstein (1958) which is most probably a translation error. German has a particular word for holiday: "Ferien", "goes on holiday" would subsequently be "in die Ferien geht". We suspect that Anscombe confused "Ferien" with "Feiern/feiern" due to their strong lexical similarity.

[3]With use *sciences of language* as an umbrella term for the following language related scientific disciplines: philosophy of language, linguistics, pragmatics, developmental psychology, psycholinguistics, antrophological linguistics, discourse analysis, conversation analysis, automatic speech processing (ASP), natural language processing (NLP), language related developmental robotics.

tempting in order to set a point of reference, a bedrock and starting point, against which we can pitch our theories. It is the basis on which we formulate our research hypotheses and how we frame the learning problem. For these reasons assumptions on universals determine on what problem we focus our efforts, what questions we ask, which kind of 'humanistic' literature we read to form the theoretical backgound of our algorithms and experimental designs, and whom we, as roboticists, potentially ask for advice in times of (theoretical) doubt.

A philosopher of language might have the luxury to suspend such assumptions, indeed we expect philosophers to be critical about any scientific assumptions, especially if they stand on a fundament as weak as is typically the case in language-related disciplines. Developmental roboticists on the other hand have to make up their mind, because these universals, thought to be essential to the capacity of 'languaging' of any human being, determine what we, as developmentalists, are allowed to design and hard-wire into our robots as opposed to what the robot is supposed to learn. Analogously those innate psychological capacities, but also a care-takers behavioural adaptations when interacting with a toddler, count as universals that are supposed to be available to human babies from birth and adhered to by there caregivers: biological and psychological givens, that constitute the necessary and sufficient elemental driving forces that enable humans and, potentially, at some point, robots to acquire this unique capacity amongst all species. In this regard universals form the basis of our acquisition or learning algorithms.

The bad news is that far more authors than only the linguists which support the nativist position make universal claims, yet many do not mark them as such. Most of the prominent theories in pragmatics such as speech act theory, Gricean politeness theory, and relevance theory have recently been identified as "universalist" (cf. Goddard 2006, Hatch 1999) and Anglocentric (Goddard 2006). Yet these approaches have never presented sufficient

empiriral support to back such grand claims.

The good news is that these premature claims for universality have been recently uncovered as such by cross-language linguists and pragmaticists, that collected empirical evidence to refute them. Furthermore these researchers, some of which now go under the name of *ethnopragmaticists* indicate that we finally seem to get some ground under our feet in terms of the question what is truly universal to *ALL* languages on our planet (Goddard 2006, Wierzbicka 1994). The author unfortunately discovered this new line of research and new methodology only upon writing up this thesis and could therefore not attempt a conversion of our negation 'types', which will be introduced later, and parts of our other terminology into their new format called *natural semantic metalanguage (NSM)*.

In order to avoid ill-formulated and ill-founded unversalist assumptions it is important to realise that these assumptions are part, implictly or explicitly, of any language-related publication that is not purely descriptive, that is, any publication that does not just report chracteristics of a particular language or a practice of linguistic interaction, but that makes any assertions beyond pure descriptions. Furthermore, via writing about language in a particular language we also make a methodological choice in terms of the "tools of thought" by recurring to, typically Anglocentric, scientific terminology or Anglocentric metaphors.

This author is a "gebranntes Kind"[4] in the sense that he believed for a long time in these false assumptions of universality of the pragmatic kind, in his particular case the universality of speech act types as proposed by Searle. As we have realized now, only cross-linguistic research is an approach that stands a chance to uncover what truly is universal. Even philosophers, the majority of whom are predominantly of Anglo-European descent, and supposedly the most general of thinkers, are prone to fall for, or make wrong

[4]This expression derives from the German proverb "Ein gebranntes Kind scheut das Feuer" which roughly translates to "Once bitten, twice shy". The literal translation would be "A burnt child dreads the fire".

claims of universality by virtue of having been raised and cultivized in an Anglo-European culture, and by virtue of using one particular language, typically their native language, as their 'tool of choice'. The fact that English is the de facto standard publication language of modern science does not really help things in this regard. Facilitating misleading ideas about universals in language-related research is the circumstance that we, as language researchers, do not only think *about* language and language-related cultural practices, but that the formulating of ideas is done *in*, and therefore influenced *by* a particular language. This is for example not the case in mathematics where the *language of the trade* is a formal, artificial language, a language nobody has been raised *in*, and which is not used in ordinary life such as in ritualistic practices (marrying a couple, christening a child etc), poetry, entertainment, or, just a for a chat in the pub - mathematics "does not celebrate".

Thus any researcher or philosopher, that thinks, writes, and talks in his or her own native language about language-related phenomana is necessarily influenced by certain, not necessarily conscious, cultural rules and what one may call concepts or metaphors in his or her very language, many of which are culture-specific and therefore not directly applicable to and available in other cultures. These rules, or cultural scripts, bias our thinking and doing in many ways that we more often than not are oblivous to. As Underhill has aptly put it: "It would seem that in language there are no 'free-thinkers'. Thought is language-bound" (Underhill 2011).

Ubiquity and importance of language for human societies Language permeates our societies and cultures on all levels from the political to the emotional. Many cultures give rise to institutions, by anchoring them in society via legally binding texts which adhere to a particular jargon or group-internal discourse with lasting and very practical effects

for the ordinary lives of the members of society. A recent historical example of such *doing with language* is the establishment of what was termed "homeland security" in the United States. The 'political erection' of the European Union, which essentially is based in a set of legal texts that constitute legally binding contracts between its member countries, did not just yield an abstract contract between societies, that is only truly acknowledged by the ones in the upper echelons of the political establishments of societies, but has very practical consequences for any member of this union. The fact, that we can just walk across country borders without showing our passports and without being asked potentially embarassing questions, the fact that we can just move to another member country and work there without having to apply for visas, i.e. without scribbling certain things on a certain kind of form, is a direct outcome of some people writing the 'right' kind of text in the 'right' manner by using the 'appropriate' language, i.e. a language acknowledged as being appropriate by a group of fellow conspecifics that we call "lawyers"[5]. The example just given is an example of how certain kinds of language use affect us, our freedom and potential actions in a top-to-bottom kind of way due to the fact that only a small group of people are endowed by the citizens of their respective countries to discuss and decide on the particular details of these foundational documents. In democratic societies the way this 'endowment' is *done* is in a bottom-up approach called election that again involves the heavy use of language, first in the form of political advertisement, propaganda, TV adverts, posters etc., all with the purpose of influencing enough people to produce the right kind of sign in the right kind of box on so-called election forms, such that 'the right kind of guy' is endowed with the right kind of power. Effectively we see here a top-down (political advertisement), bottom-up (filling out the election form), top-down use of language.

[5]Of course this is a very simplistic depiction of the actual process. The ideas that were rigidified in these contractual texts for example had first 'to be sold' to the members the participating countries. Yet this work was again accomplished by use of language. Try to subtract all linguistic doing from the political practice and odds are that you will end up empty-handed.

Yet this is only one rather abstract 'level' of language *doings*. Other ceremonial uses of language that may seem 'closer to home' and therefore a few levels 'below' the just described political doings by speaking or writing and which most members of western societies participate in are *christenings*, *weddings*, *obituaries*, and *funerals*, typically in this order[6]. Much of our social order is established via the use of language by the 'right' kind of people in the 'right' kind of position. These people are endowed with these positions by means of language use of other conspecifics: we respect policemen as such and behave accordingly[7] in their presence because they have been *made* policemen via a ceremonial practice executed via speaking, possibly rigidified via a *written* contract by other humans. Teachers, professors, CEOs, and chimney sweepers, have these kind of professional roles because some form of employment contract makes them so, which is formulated and written in language sometimes in combination with some kind of ceremony that involves a particular kind of speech. Yet another level 'down', and typically considered less ceremonial, are other language doings such as *swearing*, *insulting*, but also *soothing*, *singing a lullaby*, *waking somebody up*, or *declaring one's love to another person* which involve what we may call *emotions*. Language doings of this kind then often result in or effect what one may call emotional states such as *feeling insulted*, *calming down* or *being calm*, *being awake* or *being in love* or *feeling loved*. These are only a few examples of what we do to and with each other via 'languaging' and are meant to show that language and language-related practices are not only *part* of most human societies but must be rather seen as *constitutive* of these societies. Our societies are much less built on bricks and cement than they are built on words and by the use of words.

It is a curious circumstance that we all to easily forget about the power of 'languag-

[6]The ordering of christening and wedding may be inversed, depending if the 'target' of the practice is oneself or if one is 'only witness' to the practice.

[7]Even if we do not behave accordingly, thereby risking arrest, we typically do so in full awareness, that we are violating a cultural norm.

ing', possibly for the same reason that we typically don't think about our heart beat. It is an essential, probably the most essential ingredient involved in the construction and maintenance of the social fabric that connects us as social beings.

Negation and logic This thesis is about negation and how it might be aquired by toddlers or robots. For logicians as computer scientists negation is most often a matter of logic, and one of the central notions of logic is truth. The notion of truth is central to modern western sciences. This is also the case for computer science whose core concepts and design principles were shaped mainly by physicists and mathematicians who, during the inception of computer science, relied on and were already used to the formalism of predicate logic. Predicate logic was developed between the late 19th and early 20th century by the likes of Frege, Russell, Peano, Moore, Wittgenstein and many more, possibly lesser known logicians (see also Ferreirós 2001). It was quickly accepted by mathematicians and physicists as the common language of their trades - trades, particularly in the case of physics, whose job it is to describe the fundamental laws of nature.

Another motivation, apart from developing a formal and sufficiently powerful 'way of writing' for the natural sciences and mathematics, in the context of formal logic, seems to have been the endeavour to find an alternative and more formally rigid way to 'translate' ordinary sentences of language, any human language, into something less ambiguous and 'messy'. In the latter case the 'content' of the sentences to be pinned down is not constrained to the domain of natural sciences. In this case also 'ordinary' sentences, written down utterances, such as "Yesterday I walked my dog" ought to be captured by the formalism. It is important to emphasize the difference between these two motivations. In case of the natural sciences the phenomenas captured and pinned down by the formalism, the sentences that describe the target phenomena are, at least ideally, independent of the one

who pins them down. So the fact that they are written down by a human is not captured by the formalism of predicate logic, simply because the observer is thought to be irrelevant to the truth or falsehood of the observed. This can only be hinted to in the meta-language, which, by definition, is not part of the formal language itself. In the case of language in general, this assumption does not hold. Utterances and sentences in everyday language are phenomenas that only occasionally have to do with absolute truth. Many utterances in ordinary language do not even have the tendency to establish truth, i.e. to describe an observed phenomenon. The tendency of ordinary language, as opposed to the formal language of natural science of not being 'representational', i.e., not *being about* establishing truth- and falsehood will be described later in greater detail in chapter 2.1.

Nevertheless it is no big surpise that formal logic lies at the very heart of computer science. This is possibly due to the fact that the field is based on mathematics and engineering, both fields in which predicate logic is the standard notation for writing down ideas or laws, or to describe systems in an unambiguous and formal way. On a technical level, that is on the engineering side, the very core of the artifacts which computer science is all about are, by design, nifty combinations of AND and OR gates, that, linked together make up the core of the majority of all computational units. The influence of formal logic does anything but end with its virtual materialization in terms of silicone chips. Theoretical computer science is possibly closer to mathematics than any other subfield of computer science and it does not surprise that most formal proofs utilize the logical notation in order to keep proofs short and unambiguous. This is what the notation was created for. In another subfield of computer science, artificial intelligence, the impact of the logical formalisms was and is equally considerable. Thus, it would be hardly surprising, if one would expect a treatise on the acquisition of negation involving robots to be mainly an exercise in logic. Yet it is not.

A very conscise, albeit somewhat cryptic, explanation why this is not so, can be given in three sentences: "No" is not "not". There is a reason why the logical operator is a not-gate and not a no-gate. And finally: "No" comes first.

Words, sometimes in isolation, sometimes in company of other words and strung together to multi-word utterances are very powerful tools which humans use to act upon each other as well as to act in communion. And as with most tools they can be used to accomplish things that are beneficial for a majority of people but they can also be used in a destructive way. A hammer can be used to build a house to give shelter, but it can also be used to kill a person. A single utterance can cause two nations to go to war with each other, but it can also be used to accomplish the maintenance of peace. Pub brawls typically start with the exchange of utterances between parties which may or may not be sympathetic towards each other at the start of the verbal exchange. Chains of events, or rather chains of actions of this type have even become an idiom in German: *ein Wort führte zum anderen*, *one word leads to another*. Often, in the case of a pub-related argument, the exchange of words leads to an exchange of fists, but sometimes they don't. The pivotal activities, which decide if an argument becomes physical or not, are typically to be found in one or more utterances being uttered: one or both parties either "pour oil on troubled waters" or "fuel the fire". And this is accomplished by uttering the right or wrong words at the right or wrong time. In case of an argument becoming violent we then often say that the situation escalated. This way of explaining events does not indicate a fundamental difference between the non-physical early stage of the argument and the physical one. The transition from speaking or shouting to hitting seems to be fluent and, at least in hindsight, hardly surprises the participant or observer. If participants of such an event are asked, why they hit another person and possibly even killed him or her the justifciation is often given in terms of provocations: "John provoked me by saying X", or "John provoked

me by calling me X", or simply by "He was asking for it". So at least in the mind of the person who justifies his physical acts with such explanations, the verbal acts of the other party seem to be causal in bringing about the physical acts of violence. In other words, the utterances and words of the provoking party are seen as full-fleged actions which, by virtue of being uttered by the provoking party, led, causally or not, to a physical action on the part of the provoked.

1.1 Motivation and Goals

The major topic of this thesis it the acquisition of linguistic negation. Our goal is to shed more light on the details of this acquisition process. We try to identify the prerequisites and the particular 'mechanics' of the acquisition process in a manner far more detailed than is typically the case in psycholinguistics or developmental psychology. The way in which we are attempting to do this is akin to other constructive approaches such as artificial life. Constructive approaches (Nehaniv et al. 1999) typically transfer ideas and results in both directions, from the natural to the constructed, and back. Thus also the goals can be formulated from either perspective. Let us mentally put 'the natural' on the left, and 'the constructed' on the right, with 'the natural' being research on language acquisition in humans, typically conducted within the fields psycholinguistics and developmental psychology. 'The constructed' on the right are attempts to enable machines, in our case robots, to acquire, learn, or understand human language. The work presented within this thesis uses methods which are typically found in developmental and/or cognitive robotics and human-robot-interaction.

Generally within this kind of research, in the left-to-right direction, in terms of the flow of ideas, i.e. the right using ideas developed by the left, the goal is to construct machines

that are capable of a conversation with human beings based on insights borrowed from the left. In other words: The goal is to create robots that are able to speak and understand what their locutor says.

In the right-to-left direction, i.e. the flow of ideas and results from the artificial to the natural, the goal is to test hypotheses that have been proposed on the left by way of implementing the necessary mechanisms in machines on the right, and feed the results back to the left.

This constructive feedback loop can be beneficial to the 'natural' side because a great amount of small detailed problems have to be solved if such ideas are to be implemented on machines. These seemingly small details are easily overlooked when concentrating solely on 'the big picture' and when only considering organisms that acquire the skill in question fully automatically. Many 'small' mechanisms that play a role within human communication might be overlooked or considered of minor importance if they don't fit neatly into the linguistic theory of choice. It is only when we are forced to implement these grand ideas in software and hardware, that some of these disregarded small mechanisms can come to haunt us. And, as we will show within this thesis, some of these 'small details' have the potential to turn the theoretical edifice upside-down and make us reconsider some claims that otherwise could have lived on happily ever after in utter ignorance of reality.

As already mentioned, the particular phenomenon under investigation within this thesis is the developmental origin of human linguistic negation. In simpler terms, we attempt to answer the question: How do children learn (to use) negation appropriately, and what are the required (cognitive/computational) mechanisms to achieve this feat?

We will in particular look at two not mutually-exclusive hypotheses which we encountered in Pea (1980). The first hypothesis, the first part of which is based on said publication, but which, within Pea (1980) is not presented as such hypothesis, goes as follows:

Hypothesis 1 Negation is acquired by children by way of parents interpreting the childrens' physical display of their negative motivational/emotional/volitional[8] states in a linguistic manner. We expect negative words to be highly prevalent within these "negative intent interpretations". Furthermore we expect these negative words to be prosodically salient, such that the child may easily pick them up. Moreover, we expect these negative intent interpretations to be produced simultaneously to the child's display of those states, at least in the majority of cases, such that the negative word may be associated with the negative motivational state.[9]

Notice that the actual hypothesis is the assertion stated within the first sentence. The subsequent three sentences contain details which are not elaborated on by the main assertion. Main assertions like the above are frequently stated in the developmental literature. The subsequent details are most often not made explicit within this literature and the relevant experiments are often not executed or analysed with sufficient detail to make such 'minor' assertions. This also means, that the main assertion could still hold even if one of the details should turn out to be false, though a refutation of any of the 'sub'-assertions would shed serious doubt on the validity of the main assertion. The second hypothesis is based on Spitz (1957) and goes as follows.

Hypothesis 2 Negation is acquired by childen by way of parents prohibiting them from doing something. Prohibitions that are performed by speaking typically contain or solely consist of negative symbols. As early forms of prohibition typically go along with corporal

[8] We will later explain, why we cannot decide on any single element of this trinity.

[9] The original formulation in Pea (1980) is: "Parents also frequently interpret these behaviors (physical means of rejection) as expressive of negation and expand them with lexical negative "no, no, don't want it." Ryan (1974) has emphasized the importance of such intent interpretations for the eventual linguistic expression of intention." (Remark in brackets added by the author)

restraint, children may associate the negative symbol with the negative emotion that ensues from their limitation of agency brought about by the parental restraint.[10]

1.2 Overview of Thesis

This thesis should come with an appended DVD which contains more than 8 hours of experimental video recordings. We strongly urge the reader to watch at least some of these videos to sensitise him- or herself to the issue. Interaction is best observed, instead of being read about. Some pointers to interesting positions within the videos are provided in chapter D.

Main Section

In the present chapter we prepare the ground for this thesis: the two investigated hypotheses on the origins of negation are introduced and motivated.

In chapter 2 we review ideas from various scientific fields that we consider important for the understanding of negation. As this thesis is situated within robotics and computer science, more time is spent on 'humanistic' ideas. In our view the problem of understanding early forms of negation is one of taking the right perspective and getting the right footing rather than finding some specific kind of algorithm.

In chapter 3 we will describe our robotic architecture that was designed and constructed based on the insights gained from a cognitive 'requirement analysis' of early negation that

[10] Original formulation in (Pea 1980, p. 178 with reference to Spitz 1957: "Spitz (1957) sees the child's uncompleted act in conjunction with the parent's negative word or gesture as a major source of the first meaning of "no" for the child. His account assumes that the child's frustrated *id* drives thereby endow the negative word and gesture with a specific affective cathexis that ensures the child's remembrance of the negative symbols. The child's first use of negation, on this view, is a result of identification with the prohibiting parent, and refusal or rejection is the first meaning since the symbol is imbued with aggressive cathexis in the unpleasurable experiences associated with its memory traces."

has been performed more than 30 years ago by Roy D. Pea. This architecture will then subsequently be put to use in a set of experiments that attempt to either support or disprove two hypotheses on the ontogenetic origins of negation.

In chapter 4 we will describe the experimental setup of both the so called *rejection* and *prohibition* experiments. Both experiments have been designed to pitch the two research hypothesis against each other.

Chapter 5 is the by far most extensive chapter of this thesis. There we perform an analysis of the speech gathered during the experiments from both, the participants and the robot. The main analysis is performed on three different levels: utterance, word, and pragmatic level. Futhermore an additional analysis on temporal relationships between non-linguistic events and participant's speech acts is performed in order to answer some questions that came up due to a rather surprising result.

In chapter 6 we discuss some new issues that were thrown up during the analysis. We also quickly summarize the contributions of this thesis, discuss its potential impact, and the future work, which we are planning to follow up with.

Appendix

Chapter B contains tables that are too long for the main part of the thesis and other additional listings connected to the analysis as well as the coding scheme used within the pragmatic analysis.

Chapter C contains a short overview of customary symbols employed in conversation analytical transcripts.

Chapter D gives a short overview of the contents of the DVD. Here we also provide some

links to particular positions within the experimental video recordings that are of interest for the purposes of this thesis. As there are alltogether 98 videos (2 videos were accidentally overwritten during the execution of the experiments), the reader is well advised to start a potential journey through the provided videos by starting at one of the positions indicated there.

> The notion of a rule is logically connected to the notion of following a rule, and the notion of following a rule is connected to the notion of making one's behaviour conform to the content of a rule because it is a rule.
>
> —*John R. Searle*

Chapter 2

A Short Review of the Sciences of Language

2.1 Doing Things by Speaking

2.1.1 Logical Limitations

As alluded to in the introduction, speaking can be a form of acting. This is not to say that at some level some utterances are not amenable to a translation into a logical predicate. Yet even for these utterances an important question is, if they are indeed sufficiently characterised by the predicate. Moreover, there are many utterances that certainly are insufficiently characterised by a logical description. Furthermore there are utterances where it seems highly construed to posit the existence of a logically structured thought that accompanies such utterances, a thought held by a speaker who engages in them. Think of utterances such as "ouch!", or even "ok". For the latter kind of utterances the construction of such a logical "representation" seems utterly unnecessary and a potential waste of cognitive processing.

Let us call the paradigm that posits that each and every utterance has a propositional representation at its core *propositional paradigm*. Assume that this paradigm posits, that every human, including small children, hold a logically structured thought in their mind every time they produce an utterance. Sometimes the logical structure of this thought is called the *propositional content* of an utterance. This is somewhat of a misnomer as the content is not thought to be attached to the utterance but is rather thought to be held in the mind of speaker at the time of speaking. The particular kind of logic to be employed as "language of thought", i.e. the particular logical flavour of this "internal language of thought" is not overly important. But it might be worth saying that there are strong indications from research on biases in human thinking that suggest that the type of logic at play would most probably not be the standard predicate logic (Levinson 1995).

It seems that the propositional content at the core of an utterance is thought to represent whatever state of affairs in the world the utterance refers to or, in other words, what the utterance *is about*. Saying that propositional contents are at the core of an utterance hints towards the existence of a periphery, something secondary, which might not be amenable to truth-functional logic. J. L. Austin developed a theory, speech act theory (Austin 1975), that investigates parts of this so-called periphery which characterizes many utterances as so-called speech acts. John Searle, his student, subsequently developed a categorization system for these speech acts (Searle 1969). Speech act theory will be sketched a bit further below. First we will give some examples of speech in (inter-)action where this so called periphery seems to become surprisingly dominant in terms of explaining the character of a spoken utterance.

Utterances where people *give reasons* in order to justify their actions are one example where predicate logic comes dangerously close to its descriptive limit. When a speaker gives a reason why she did this or that, i.e. unveils her motivations behind a certain deed, this

act is by its very nature subjective. A curious property of reasons is that a hearer has no problem whatsoever in accepting a reason as a reason given by some other speaker, which he himself would never give as a reason for the same deed. Assume somebody explains in court that he killed his wife for cheating on him. In this case the accused states the fact or at least his belief that she cheated on him, as the reason for his deed of killing her. We might not accept this reason as sufficient to justify the deed (murder/manslaughter). But this does not imply that we don't take this to be a reason of the speaker. By saying "But this is not a reason" (German: "Das ist doch kein Grund!"), we do not say that we don't think that it is not a reason for the speaker. What we say is that, if we were in the place of the speaker, we would not consider this reason to be sufficient to justify the deed, possibly implying our adherence to a higher moral standard compared to the speaker. We say that the given reason is not a reason *for us*, but we don't say anything about the reason being or not being a reason for the speaker. We can at most speculate what would or would not count as a reason for the speaker. Absolute truth about this 'subjective truth' of the reason being a reason for the speaker cannot possibly be established in such a case by any person other than the speaker due to the lack of access to the internal state of the speaker and his ways of reasoning. Predicate logic is not designed to handle anything else than absolute, that is speaker-independent truth.

In an attempt to rescue the descriptive power of the notion of truth for natural language, this notion would have to be modified considerably. One could, for example, try to establish a notion of subjective, or speaker-dependent truth in order to accomodate the fact that different speakers often do not agree on what is and what is not the case. As in the previous example, a reason might be a proper reason for one person, but does not even come close being a reason for justifying the same action for another person. Certain forms of modal logic, which are extensions of predicate logic, ought to capture 'subjective truth'

by introducing operators for *believe*, thereby introducing an agent-centric perspective into the formalism. These kinds of logic have no problem with formalizing truth-functional disagreements between various speakers by attributing different beliefs to different speakers. Thus speaker-dependent truth can be handled by extensions of predicate logic. Yet, from a logical perspective, two or more speakers sometimes do much worse things than just uttering things that are logically incoherent when engaging in a conversation. These 'speech acts' are worse in terms of the degree to which they could be considered truth-functional or representational.

One example of such an 'illogical' deed was called performative by Austin. "Close the window!" for example cannot be sufficiently characterised by 'beliefs' no matter how weak of a truth they ought to be. The reason for this insufficiency is that performatives are not 'about' something else being or not being the case. Even worse for any truth-functional account of language, they only marginally involve the speaker's beliefs if we judge them for their communicative success. Their degree of 'being about something that is the case', is close to zero.

2.1.2 Speech Act Theory

In this subsection we quickly introduce speech act theory by examples. The particular 'implementation' of the theory is not of great importance to our purposes and we refer to Levinson (1983) for a good introduction into the theory and an excellent formal proof of some of its shortcomings. The purpose of this section is merely to further sensitise the reader to the non-logical aspects of language use in preparation for our taxonomy which will be introduced in chapter 5.3. We will generally adhere to Austin's (1975) account of speech acts as outlined by Levinson (1983) unless we say differently.

Performatives vs. constatives *Performatives* were the first types of speech act that caught Austin's attention, who is commonly thought to be the father of speech act theory (*SAT*). At first he contrasted performatives with *constatives*, the latter of which basically subsume statements and assertions, bearers of truth. Conversely, what performatives 'are really about' is to *bring about* a future state of affairs, 'fact creators' so to say, or the manipulation of the addressees' state of mind. In other words, these utterances, or speech acts, are full-fledged actions. As performatives, such as "Could you please close the window", a request, can generally not be evaluated in terms of truth, Austin introduced the notion of *felicity*. The qualification of a speech act as *(non-)felicitous*, i.e. the (non-)successful performance of a speech act, then may be seen as a more general qualification of (the act of producing) an utterance and replacement for *true* and *false*. There have been attempts to 'loosen up' semantics in order to accommodate these 'non-representational' acts of speech in order to circumvent the need for a separate (speech act) theory but none of them succeeded according to Levinson (1983). Later Austin evolved his taxonomy by giving up the strict separation between performatives and constatives, to make the latter a specific case amongst many other cases or types of speech acts. Due to space considerations we refer to Austin (1975) and Levinson (1983) for further details of Austin's taxonomic development.

Austin further noticed that the performance of certain speech acts appears to hinge on the existence of *conventional procedures*. In order to christen a baby, for example, one has to follow certain socially established rules, which involve a certain location (registrar's office or church), a certain role of the baptist (priest), and the use of certain (ceremonial) utterances. Certain types of speech acts further require the addressee to acknowledge or ratify their performance, as is the case when betting somebody as in "I bet you five pounds that Arsenal will win tomorrow". Without the addressee acknowledging the act with something like "you're on", the bettor cannot be said to have actually betted. Yet other

types of speech acts such as promises require the promising party to sincerely believe that it will stick to the promise, and condolences require the speaker to feel actually sorry for the addressee, if they ought to be felictous and not just empty formulae. Austin developed a set of criteria that ought to capture the kind of conditions that have to be met in order for a speech act to be considered as having been performed and, further, as having been performed felicitously.

Felicity conditions Firstly (A), there has to be a conventional procedure with a conventional effect, and the speaker, if this procedure requires, may have to hold a particular role and may have to be in a particular situation[1]. Secondly (B), this conventional procedure has to be executed correctly and completely. And thirdly (C), the speaker may have to have the required thoughts, feelings, or intentions, as specified by the procedure, and, further, may have to follow-up with certain actions or a certain conduct, if the procedure requires.

Non-acting vs. acting unsucessfully Clearly, any of these criteria may not be met. The outcome of not meeting one or more of these criteria may have two different kinds of failure as consequence. Austin distinguished between the non-performance or the non-coming-off of a speech act, which is the case if any of the criteria listed under (A) or (B) are not met, and a speech act beeing performed unfelicitiously, which would be caused by a failure to meet any of the conditions listed under point (C). An example for the first case, the non-performance of a speech act, is if a random person on the street approaches you and says "I hereby sentence you to 5 years in prison". This is a non-sentence due to

[1] In the case of christening a child, it is not sufficient that there is a baptist and some random child. The parents of the child actually must want the child to be baptized at that very point in time, must typically be present, and they typically have to be members of or believers in the respective religion. This is what is meant with "particular situation" here.

the random person not being a judge. Even if this random person happens to be a judge, it is not the right place (court), nor is it the right situation (trial) for a sentence to "take off". In these cases the speech act is thus not only not felicitous, the act was indeed not carried out at all or *misfired* in Austin's jargon. An example of the second case, a speech act being performed unfelicitiously, would be if I promise you that I will do something, for example attending your birthday party, without having the intention of doing so. In this case, I did indeed promise, but the promise in not sincere and therefore unfelicitous.

So it is perfectly possible that a speech act was performed, but it may not be considered *felicitous*, due to one or several criteria being violated. If somebody tells you to close the window and you follow the request, it doesn't matter if the speaker believed that you would do it or not in order for the request to be a request. Neither does it matter if the speaker actually believes that the window is indeed open or not. If you turn around to close it and realise that the window is already closed, the request was successful in being a request. It was not *felicitous* though if the person, asking you to close it, did not believe that the window was open in the first place and just used this utterance to annoy you. In this case the speaker would have acted insincerely, which is a violation of the criteria listed under point (C). If the speaker actually believed that the window is open, but it is not, the speaker acted based on a false belief. The latter kind of failure is not covered by Austin's criteria.

As can be seen from this list of criteria, only required beliefs can be analysed in or reduced to terms of truth- or falsehood. A request as the aforementioned may be caused by my false belief of the window being open, but this being the case does not render the act of requesting a non-act. Compare: Somebody throws a snowball at you, planning to hit you, based on the belief that you were the culprit who threw the previous snowball which hit him. In this case neither the thrower's belief of you being the culprit (false belief), nor

his belief, that he will indeed hit you bare any relevance on the decision as to whether he threw the snowball. The act of throwing was performed even if the thrower does not hit the target. Acting on a false belief, is acting nevertheless. And not achieving one's goal by an action does not render an action a non-action. The distinction between successfully throwing a ball in terms of actually having thrown a ball, not having dropped it, on the one hand, and successfully throwing a ball in terms of having hit the target, on the other hand, was formalised in speech act theory.

Locutionary vs. perlocutionary acts This distinction is described by the notions of *illocutionary* and *perlocutionary* acts. More precisely, these two kind of acts, terminologically somewhat misleading, don't describe different actions, but rather characterize different levels of the same action[2]. Formal success criteria were only defined for the illocutionary level and are the criteria we grouped into the three categories A, B, and C above. Generally, speech acts are identified with the illocutionary act.

The *illocutionary act* is the act performed by producing a (grammatically and semantically correct) utterance *and* by virtue of the illocutionary force that comes with this type of utterance. For example a request is only a request in a particular language if it has a certain grammatical, or prosodical structure, and is issued in the right kind of circumstances. It is only a request instead of being a random utterance because there is a norm (or conventional procedure) that is recognized by the the speakers of this language, and the illocutionary force of "requesting" is bound up and reliant upon the conventional procedure in order for the act to be successful. This norm then leads to there being the illocutionary force "request" that 'comes with' the utterance. Other examples for illocutionary acts are

[2]There is a third, "underlying" level, the so called *locutionary act*, the act of having said something unambiguously, which roughly maps the linguistic and semantic level. We only mention this for completeness here.

promising something, *sentencing somebody*, but also *making a statement*. The success of the illocutionary act, as opposed to the success of the perlocutionary act, does not depend on local circumstances other than the ones specified by the conventional procedure. Its successful performance therefore depends mainly on the speaker. The only requirement for an illocutionary act to be successful, that does not rest with the speaker, is the so-called *uptake* of the act. This means that the speaker must ensure that both the content of the utterance *and* the force of the utterance are understood by the addressee. Furthermore, some speech acts require some form of ratification on part of the addressee, such as the abovementioned ratification of a bet.

The *perlocutionary act* or *effects* of a speech act are the effect that the speaker tries to achieve by uttering the utterance with the particular illocutionary force. As opposed to the illocutionary act, the perlocutionary act very much depends on local circumstances within which it is performed and "is therefore not conventionally achieved" just by the production of the utterance according to the norm (Levinson 1983, p. 237). Apart from the effect, that the speaker intended the utterance to have upon the addressee(s), all other (side-)effects, planned or unplanned, that the utterance may have on the audience also count to the perlocutionary effects. Examples of perlocutionary acts are *forcing somebody to do something*, the potential outcome of the illocutionary act *giving an order*, or *making somebody believe that one is commited to do something*, which may or may not be achieved by *promising to do something*, an illocutionary act.

It should also be noticed that the border between illocutionary and perlocutionary acts is not as neat as it may seem at this point (cf. Levinson 1983). Especially the circumstance that a speaker has to secure the uptake of an utterance and ensure potentially necessary ratifications are both deeds that are not totally under the speaker's control and are clearly not exclusively an issue of sticking to some conventions. This blurs the border between

the illocutionary and the perlocutionary and Levinson (1983) concludes that "there seems to be no clear reason why what is a perlocution in one culture may not be an illocution in another" (p. 241).

After Austin, speech act theory was further developed by John R. Searle, one of his students (cf. Searle 1969). At this point it shall suffice to say, that Searle put the theory on a slightly different footing and introduced a pseudo-axiomatic categorisation of speech acts, which, according to Levinson (1983), "is a disappointment in that it lacks a principled basis" and "contrary to Searle's claims, is not even built in any systematic way on felicity conditions" (p. 240). Nevertheless Searle's version of speech act theory may be the most known and popular one. Yet, possibly due to Searle's unsatisfying categorization, several competing taxonomies exist now, and we refer to Levinson (1983) for further links.

With children that have not yet acquired either the necessary social knowledge or the required skills in terms of motor control, it has been observed that they use different kinds of "manual" gestures such as reaching and pointing gestures in order to signal their communicative intention, which might be early forms of what later become speech acts (Clark 2009, ch. 4). Yet within speech act theory itself developmental considerations are at best rare. There, the invoked examples of acting by speaking typically assume that the speaker is a human with full linguistic capabilities and equally full awareness of the social norms or conventional procedures in place. Nevertheless speech act theory has been adopted by some researchers of early child language development, but these accounts typically diverge considerably from the original theory due to conversational properties of actual talk which Austin and Searle have either overlooked or ignored when constructing the theory. We will pick up developmental perspectives on speech act theory in section 2.2.2.

Here is another example to illustrate the power of acting via speaking: If a judge sen-

tences somebody to 10 years in prison, it doesn't matter, if he believes in the accused's guilt or not in order for the sentence to be at least temporarily in place, i.e. for the act to be publicly recognised as an act of sentencing. Even in the very constructed case, where the judge suffers from a sudden bout of some mental disease, and, at the time of sentencing, forgets that he is a judge, forgets who the accused is, and possibly who anybody else is, and just so happens to say the appropriate words "I hereby sentence you ..." at the time slot in the court proceedings reserved for the sentence, this sentence would be in place. It would be in place by virtue of him being a judge and by him uttering the 'right' words at the right place and time. In speech act theory and most probably also in law, this sentence would be considered "insincere", but it would be a sentence nevertheless. The sentence could probably be easily reversed afterwards, if the judge, let's say, jumps out of the window after having sentenced the accused, and thereby makes the authorities realise that he was 'out of his mind' when he said the fateful words. But at least for a few minutes, until the appropriate measures are taken to remedy the juridical error, this sentence would be in place, no matter what beliefs the judge or anybody else held at the time of sentencing and no matter what else was the case in the world.

On speech acts and propositional content Based on the outline in the previous paragraphs of speaking qua acting one might assert that for example performatives, *christening somebody*, *sentencing somebody*, or *ordering somebody to do something*, have more similarities with the act of hammering a nail into the wall, than with the exchange of information about something already being the case in the world. Nevertheless propositional content is still an important notion in Searle's version of SAT, and there, it seems, the illocutionary force is seen as some kind of wrapper or add-on to the propositional content.

This is also indicated by his formal notation for utterances, $F(p)$, where F denotes the force of an utterance, and p denotes its propositional content (Searle 1969). In Austin's account of speech acts the notion of propositional content features far less as compared to Searle's account. Indeed, when searching through his main publication "How To Do Things With Words" (Austin 1975), we find propositions only mentioned at three different places, one of which is a discussion of logical notions of entailment, implication, and presupposition. At the other two places, Austin does not appear to be overly convinced of their importance for the capacity of acting via speaking. There we find that sentences (or propositions) appear to be logical constructions out of speech acts (p. 20), which we may interpret as the act being the primary aspect of an utterance and the proposition being secondary. Later on in the book Austin seems to become even more suspicious of the very notion of propositions when he says "... in order to explain what can go wrong with statements we cannot concentrate on the proposition involved (*whatever that is*) as has been done traditionally" (Austin 1975, p. 52, emphasis added). And albeit Searle seems to be rather commited to propositional content being an elementary ingredient of speech acts, he also admits that "(o)f course not all illocutionary acts have a propositional content, for example, an utterance of "Hurrah" does not, nor does "Ouch" "(Searle 1969, p. 30). He later adds "Hello" to his list of examples of proposition-less speech acts.

In the context of the importance of propositional content on the ability to act via speaking, the analogy with an action that is not performed via speaking, might cast some light on the issue as to whether such an act intrinsically requires any form of propositional content.

What would the propositional content of the action of hammering a nail into a wall be? The most likely candidate seems to be the proposition "I am hammering the nail into the wall", which one presumably ought to have in mind each time it is performed - similar

to a speech act that ought to 'carry' propositional content each time it is performed. The problem with the latter is, that the proposition would only be required to be in one's mind during the act of hammering, due to the claim that the propositional content is part of the action. The outcome would be a vacuously true proposition in the best case, as the proposition would always be true when performing the action. The proposition would only be false, if the hammerer did not succeed to hammer by, for example, losing hold of the hammer before hitting the nail. Yet in this case the proposition still does not serve any function in terms of helping or being a requirement for the action to succeed. Other non-propositional corrective aids, a teacher assisting the hand or the like, might be efficacious in improving the success rate of the action. The hypothetical core proposition is quite evidently of no use in guiding the action. Of course this comparison uses an extreme case of 'pure bodily intelligence', but consider a simple "Hello", which might be replaced with a simple gesture, and the comparison is not that far off. As Searle admits on the rare occasions mentioned above, there is no intrinsic need for propositional content in order to act via speaking.

Possibly of not much concern to the philosopher, but perhaps important for 'engineers of a mind', instantiating a proposition for such a simple act as hammering a nail into the wall or greeting somebody, would be outrageously inefficient. There is no benefit in instantiating a propositional representation, which is trivially true, taking up space in the (attentional) memory and possibly starting a logical inference machinery that would consequently idle most of the time. This is not the same as saying that the agent should not be aware of the action being executed. In technical terms there are far more efficient, and minimal ways of 'consciously'[3] keeping track of what one is doing than invoking the whole logical and inferential machinery[4], especially if the latter would subsequently idle along for

[3]'consciously' meaning here, that the representation is loaded into the working memory

[4]In computers, when using logical representations one has to start an inference machinery, that is

most of the time. Considering a logical representation (and the inference mechanism that typically comes with it) only makes sense, if the agent wants to communicate what he is doing or, possibly, in order to understand what another agent said, that he would be doing. Where this is not the case it appears to be unnecessary to instantiate such representations in order to "just do things in speaking". This is the distinction that can be drawn in speech act theory between speaking qua stating facts and speaking qua acting.

Acting with voice beyond speech acts There are other ways in which we act via speaking which are not explicitly covered by speech act theory, possibly due to a low degree of conventionality, or because these acts may be considered side-effects of an utterance (perlocutionary and non-intentional effects). These may be efficacious more on a biological level without the need to invoke societal conventions. These biological effects of humanly produced acoustic waves are not limited to human speech though. Dogs employ sound waves to deter a potential aggressor by barking, and human babies call for attention by crying. The physical properties of these sound waves can have considerable effects on the recipient[5]. And a baby's cries may cause on a psychological level more distress than a jack-hammer emitting sounds on the same decibel level, because the human auditory system might be more sensitive to frequency bands typical to the human voice, than to those of a jack-hammer.

We may not conveive of these activities as full-fleged, intentional actions, but they certainly have properties that are also important in conventional speech such as the intensity

able to 'make sense of' and process the representation, for example some implementation of Prolog. This author is not aware of any technical implementation of logical representations independent of an inference mechanism and it is not clear why one would want to have that. The whole point of having predicates is to draw inferences from them.

[5]A baby's cry can reach 110 decibels which is roughly equivalent to the sound intensity of a car horn or a rock concert. Exposure to this level of 'noise' will cause damage to a person's hearing capacity after 1.5 minutes according to Common Environmental Noise Levels (2013). Thus, babies are not that harmless after all.

or intonational contour of the acoustic signal, which may become conventionalized in adult speech as so called IFIDs (illocutionary force indicating devices). Intonation contour, together with other indicators such as word order, punctuation, or performative verbs, was proposed as an IFID (Fotion 1975), i.e. as indicator as to which kind of illocutionary force is present with a particular utterance.

Utterances can be abusive or soothing, they can have a direct impact on the well-being of the person that is 'hit' by them. Non-lexical properties of speech such as prosodic contour and energy of the sound wave seem to play an important role in this context. It is for example questionable if one can verbally abuse a person in the long run, without producing one's abusive utterances with a high energy accoustic signal and without giving the abuses the 'appropriate' intonation contour. It is not entirely clear if this may be done by virtue of appealing to conventional procedures or if this may be so because of the impact on a biological level. The circumstance that babies react to some of the acoustic properties, hints towards the effect being non-conventional and biological.

Speech acts beyond philosophy and linguistics In order to demonstrate that the existence of speech acts is acknowledged not only by philosophers and linguists, but that this capacity of human speech is indeed part of our everyday life, this paragraph shall briefly demonstrate their existence outside of the latter academic realms.

The power and efficacy of some forms of verbal conduct is nowadays acknowledged by penal codes or legal acts in many countries. In the jargon of modern law malevolent verbal conduct which has a detrimental effect on the addressee's well-being has been given the label of *verbal abuse*. Bullying or stalking are other acts or forms of conduct which are typically at least partially accomplished by the malevolent use of words. Stalking was only recently added to the penal codes or codified in legal acts of several countries

and it is illuminating to analyse the precise wording of those texts. The Protection from Harassment Act 1997 (1997) defines "actions of harassment" in section 8, the section for Scotland[6], as follows:

> (2) An actual or apprehended breach of subsection (1) may be the subject of a claim in civil proceedings by the person who is or may be the victim of the course of conduct in question; and any such claim shall be known as an action of harassment.
>
> (3) For the purposes of this section –
>
> "conduct" includes speech;
>
> "harassment" of a person includes causing the person alarm or distress; and
>
> a course of conduct must involve conduct on at least two occasions.

This legal act acknowledges that speech, as one of many possible forms of conduct, has the capacity to victimize persons and that such forms of conduct are punishable within the jurisdiction of the United Kingdom. In other words it acknowledges the capacity of speech to cause "the person alarm or distress", i.e. the capacity to manipulate the psychological state of the recipient into an adverse psychological condition against his or her own will.

Thus, not only laymen, when trying to justify certain actions (or 'conducts' in legal jargon) sometimes refer to speech as being causal in povoking distress. Even in formal legal texts as the one cited above this capacity is explicitly acknowledged.

Limitations of speech act theory

Possibly the major limitation of *SAT* is its exclusive focus on single utterances qua speech acts. Already Fotion (1975) noticed that the illocutionary force can be detached from the

[6]The wording of the equivalent section for the rest of the UK was more convoluted, leading to the section for Scotland being cited.

utterance that it is supposedly part of. The first example, that he gives, is not necessarily damaging to the theory:

> Speaker 1: "You had better get a move-on"
>
> Speaker 2: "Are you threatening me?"
>
> Speaker 1: "Yes"

Here, the "illocutionary force is tacked on later (3rd move) and is not literally part of the original speech act (1st move)." (Fotion 1975, comments in brackets added by us). Because the third move is only necessary because either speaker 1 spoke ambigously or because speaker 2 failed to secure the uptake, one can count this as exception - the force was supposed to be part of the first utterance but things went wrong.

More damaging to Searle's SAT are cases which are clearly not repairs but absolutely valid conversational moves. The IFID can, for example, be specified before the utterance, 'the propositional content', to which it supposedly belongs, as in "Here is what I promise", followed by what is promised, or as in "Here are your orders for the week", followed by the orders. Thus, Fotion draws the conclusion, that "language ... is not accurately analyzed in terms of this or that isolated speech act. Instead, it is to be seen as a language which is understood more completely only when the analyst take into account the flow of language - use on the activity level - so that what was said a moment or two ago is viewed as having direct logical bearing upon what is being said now and will be said later." (Fotion 1975) This direction of criticism, coming from within the philosophical community, is very similar to the criticism, brought forward against *SAT* from a substantially different academic field: conversation analysis.

2.1.3 Conversation Analysis

Conversation analysis (*CA*) grew out of the studies of Harvey Sacks on the organisation of ordinary talk. Methodologically *CA* is diametrically opposed to SAT in the sense that it is entirely data-driven, based on recordings of *talk-in-interaction* as it actually occurs. As indicated by the notion of *talk-in-interaction*, *CA* does not exclusively focus on language per se, but rather on the particular ways that social activities are organized interactively (Hutchby and Wooffitt 1999). Yet utterances, as part of conversations, are naturally an important part of this interactional work. Contrastingly, speech act theory is based on, possibly witty, informative, but nevertheless made-up, remembered, or imagined examples instead of transcripts of actual "language at work" (cf. Wittgenstein 1984, PI 132)[7].

Apart from this methodological difference, conversation analysis, as the name indicates, extends the analytical scope to the level of the conversation. That means, that its main focus is on the sequential organisation of talk, the way that speakers organise their turns depending of their own and other speakers' previous turns. This, naturally, is also in opposition to the SAT approach of focusing on the atomic elements of a conversation in isolation.

Two notions, that are central topics within conversation analysis are important within this thesis. The first one is the notion of *adjacency pairs*, a central criterion in our taxonomy of negation types, which will be introduced in section 5.3.2, and the second one is the notion of *turn-taking*, which any treatment of human face-to-face communication cannot really do without.

[7]Levinson (1983) very aptly expresses this point, while comparing CA to discourse analysis. He says that in CA "the emphasis is on what can actually be found to occur, not on what one would guess would be odd (or acceptable) if it were to do so." (p. 287)

On rules in conversation analysis

Before turning to the abovementioned notions, we will quickly discuss the nature of conversational rules, as they are investigated and extracted from talk-in-interaction by conversation analysts.

Rules, as for example the turn-taking rules, presented in the section on turn-taking below, are normative rules. This means, that they are not seen as being causal for, constitutive of, or external to the talk. Participants of the talk typically adhere to these rules in order to accomplish the talk, not because of any hypothetical law-like nature of these rules[8]. Thus, these rules do not describe, *how* participants accomplish the turn-allocation, in terms of necessary and sufficient cognitive abilities. This also means, that these rules cannot be directly implemented by an 'engineer of a conversational mind'. Rather conversely, such an engineer would have to find ways and means to enable a conversational mind to orient towards these rules if any human-like conversational capability is to be achieved. Conversation analyses showed that participants of talk orient towards these rules, but not how participants achieve this orienting in terms of underlying mechanisms. These normative rules then constitute resources for participants to orient towards, and further allow a description of talk-in-interaction as "abstract, structural phenomenon, without going against the idea that participants are knowledgable agents who are not somehow 'caused' to act by that structure but actively use it to accomplish particular communicative actions" (Hutchby and Wooffitt 1999, p. 141).

[8]Historically, there is a well-documented 'clash of paradigms' between John R. Searle, Emanuel A. Schegloff and others on the nature of rules, whose textual artefacts were (at least partially) collected and published under Searle et al. (1992)

Turn-taking

A central notion within *CA* is the notion of *turn-taking*. This notion is based on the observation that in *mundane conversations*[9], i.e. formats of talk, in which turn form, turn content, and turn length can be varied freely by the speakers, there is a tendency that one speaker talks at any given time. Furthermore, it can be observed that there are usually few gaps between turns (Hutchby and Wooffitt 1999). A central topic of *CA* then is how this turn-taking is organised and accomplished, but also, "what participants take it they are actually doing in their talk" (Hutchby and Wooffitt 1999). Other important issues that have been investigated within *CA*, are how overlapping talk is managed, and how conversational repair is accomplished. For more information on the latter issues, we refer to (Hutchby and Wooffitt 1999, ten Have 2007).

The turn-taking model, as developed by Sacks et al. (1974), consists of two components, the *turn-construction* component, and the *turn-distribution* component (Hutchby and Wooffitt 1999).

Turns can be regarded as consisting of units, turn-construction units, that roughly correspond to linguistic categories such as words, clauses, phrases, and sentences. An important feature of these turn-construction units is that they are *projectable*. This means that a participant of a given talk can during the production of a turn-construction unit identify the kind of the unit. Moreover, he or she can estimate when this unit is likely to end. These (projected) boundaries of turn-construction units are called *transition-relevance places*, and afford participants of the talk the possibility of a transition to another speaker.

[9] *Non-mundane* conversations are ceremonial or institutionalised kinds of talk, where talk underlies more rigid contraints in terms of what, how much, and how things can be said. During baptism, for example, the baptist follows a rule-set which contrains his or her possibilites of what can be said. In western societies it would be considered inappropriate, if the latter would start the "core-process" with something like "I hereby name you", just to change the topic, after having said "Hereby", to speak about yesterday's football game.

Transition-relevance places then connect the turn-construction component of the model to the turn-distribution component. The latter descibes how turns are allocated amongst participants of the talk. The normative rule set as proposed by Sacks et al. (1974), that regulates turn-allocation, looks exceedingly simple:

> Rule 1 (a) If the current speaker has identified, or selected, a particular next speaker, then that speaker should take a turn at that place.
>
> (b) If no such selection has been made, then any next speaker may (but need not) self-select at that point. If self-selection occurs, then first speaker has the right to the turn.
>
> (c) If no next speaker has been selected, then alternatively the current speaker may, but need not, continue talking with another turn-constructional unit, unless another speaker gains the right to the turn.
>
> Rule 2 Whichever option has operated, then rules 1a-c come into play again for the next transition-relevance place.
>
> (Hutchby and Wooffitt 1999, pp. 49-50)

Notice that this rule set cannot be used to predict the serial allocation of turns such as "participant A will produce 5 turns, followed by 3 turns by participant B". This is so, because different participants can orient towards different sub-rules of the set at any transition-relevance place. Speaker A could for example attempt to take the next turn by orienting towards rule 1b, while speaker B, the active speaker, may orient towards rule 1c, which allows him to continue to speak. In such a case repair mechanisms are needed in order to remedy the ensuing trouble (see Hutchby and Wooffitt 1999, p. 57 ff. for a short summary of conversational repair mechanisms).

Adjacency Pairs

An important notion, that was developed in *CA*, and which will become important later for our taxonomy of negation types, is the notion of an *adjacency pair*. Adjacency pairs are ordered pairs of utterances that consist of a first pair-part and a second pair-part that are distributed across two different speakers (Hutchby and Wooffitt 1999). Furthermore, certain first pair-parts require one particular or a range of certain second pair-parts. In the parlance of *CA* the production of a first pair-part by a speaker A *makes relevant* the production of a second pair-part by a speaker B, for which the latter is subsequently accountable. Examples of adjacency pairs are *invitation - response*, *greeting - greeting*, or *question - answer*. Adjacency pairs are an important mechanism with which participants establish sequential order in talk-in-interaction. This does not necessarily mean though, that the second pair-part must be immediately produced in the follow-up turn as the following example of a so called *insertion sequence* shows:

A: May I have a bottle of Mitch?	((Q1))
B: Are you twenty one?	((Q2))
A: No	((A2))
B: No	((A1))

(Merritt 1976, p. 333: quoted in Levinson 1983, p. 304)

Here B, instead of answering A's question immediately, produces another first pair-part of a question-answer pair, which A answers on the next turn. Nevertheless B shows by producing the answer A1, the relevant second pair-part to Q1, in the subsequent turn, that he is still orienting to A's question. Thus adjacency pairs do not only induce a certain order, they are also used by participants to display their mutual understanding:

> What two utterances, produced by different speakers, can do that one utterance cannot do is: by an adjacently positioned second, a speaker can show that he understood what a prior aimed at, and that he is willing to go along with that. Also, by virtue of the occurrence of an adjacently produced second, the doer of a first can see that what he intended was indeed understood, and that it was or was not accepted. Also, of course a second can assert his failure to understand, or *disagreement*, and inspection of a second by a first can allow the first speaker to see that while the second thought he understood, indeed he misunderstood.
>
> (Schegloff and Sacks 1973, p. 296: quoted in Hutchby and Wooffitt 1999, p. 41, emphasis added)

Pauses/Silences Pauses are generally important elements of conversations and have been observed to be regarded by participants as "someone's silences" (ten Have 2007, p. 19). As opposed to silences within a speaker's turn, or silences between different speakers' turns outside of adjacency pairs ("between-turn pauses"), sufficiently long pauses after the production of a first pair-part of such pairs are noticeable and accountable. Together with non-deliveries of a second pair-parts that are not preceded by pauses, they constitute so called *noticeable absences* (Schegloff 1968). These noticeable absences can be meaningful in themselves for one or all speakers, yet they are meaningful only by virtue of their position within the conversation. Following is an example of a noticeable absence.

1 C: So I was wondering would you be in your office on Monday
(.) by any chance?
2 (2.0)
3 C: Probably not
4 R: Hmm yes=

5 C: =You would?

6 R: Ya

7 C: So if we came by could you give us ten minutes of your time?

(Levinson 1983, p. 320)

The pause 'produced' by R at position 2 is evidently taken by C to be a response to this question, and it is taken as a 'no'. Thus, R's non-reply was indeed taken to be a conversational move by R, as he makes clear at position 3. This shows not only that words or utterances are meaningful within a conversation, but also non-words and non-utterances can be meaningful and even take the place of words. This means that meaning in a conversation is certainly not bound to words and utterances but to some structure that is only indirectly impacted by the words and utterances, which are said, but also by words and utterances that are not said. Thus, as this short snippet of a conversation shows, there is a strong sense of some form of inferential mechanism at work, that draws on normative expectations. But, first of all, this mechanism may not be active with equal strength at any point of a conversation. The example above is a particularly strong case of one participant's (R) follow-up action being monitored by his or her interlocutor (C), due to the fact that C uttered a first pair-part of an adjacency pair. Adjacency pairs carry above-average normative expectations with regards to R's subsequent action, in this case the production of an answer. Moreover this hypothetical inferential mechanism is most probably far less concerned with state of affairs in the non-human world than it is concerned with human action and therefore probably far less 'logical' as one might expect from an inferential mechanism - at least from a computer science perspective. Sadly this mechanism is not very well researched and probably part of what Levinson calls "the human interaction engine" (Levinson 2006).

On a technical side-note, we would like to mention that Jefferson (1989) indicates that a pause length of one second appears to be a relevant threshold within conversations.

2.2 Language Acquisition in Humans

Most modern theories of language development appear to be largely "representationalist" in the sense that the main task of the child is presented as mapping the words to categories of physical objects or elements in the world, a labelling or naming task. Hirsh-Pasek and Golinkoff (2012) summarise the toddler's job as follows: "We have seen that language learning can be distilled into three main tasks: (1) finding the units of speech (segmentation) that will become the sounds, words, phrases, and sentences; (2) finding the units in the world (objects, actions, and events) that will be labeleled by language; and (3) forming mappings between elements of the world and words." An important constraint in this context that aids children with their labelling task, is the so called whole-object-constraint (Markman 1990). Under this constraint, children seem to assume that a novel word, all other things being equal, refers to a whole object, and not to its colour, its texture, nor to its spatial relationship to other objects.

But what about negation? A single *no* clearly cannot be mapped to any category of objects. It should also be noted that in terms of experimental support, the whole-object-constraint has been shown to apply to children of one-and-a-half years, yet most empiririca evidence was accumulated for three and four year olds (Clark 2009). The simple *no* on the other hand is typically produced before the first object words are. In the case of American toddlers 50% of them produce *no* roughly after *daddy*, *mommy*, *hi*, *bye*, *uh oh*, and *dog* (Fenson et al. 1994, p. 93). According to their caretakers their, i.e. the caretakers', *no* is perceived to be understood as early as at eight months of age (Fenson et al. 1994, p. 92).

Ryan (1974) who provides us with an important hint about a potential developmental source of negation (cf. hypothesis 1 in section 1.1), criticises accounts similar to the one of Hirsh-Pasek and Golinkoff (2012) above for their tendency to postulate a transition from a "predominantly expressive or emotional use of words to their factual or descriptive use" and a frequently encountered disregard for communicative functions of early utterances. The problem with latter transition, according to Ryan, is that it "does not make sense in that there are no possible empirical criteria on which such a distinction could be based."

Historically, while Vygotsky still speaks about emotion in the context of language development, apparently supporting Meumann's idea " that the first words are also purely affective, expressing feelings and emotions" and "are devoid of objective meaning, reflecting, like an animal's "language" (Vygotsky 1986, p. 93 refering to Meumann 1902), affect and emotion are not even listed in the index of some modern accounts of language acquistion such as Clark (2009) any more.

Thus, representationalist theories of language acquisition are not very helpful in terms of explaining the emergence of early forms of negation due to their almost exclusive focus on words that ought to refer to objects, object categories, or physical events and actions, and the cognitive mechanisms that are involved in accomplishing the correct word-to-concept mapping.

2.2.1 Early Words

"While infants vocabularies contain a variety of words, the early vocabulary of young learners has often been characterized as biased towards nouns. Nouns form the majority of children's early receptive and productive vocabulary and are typically acquired earlier than other word classes in many languages" (Poulin-Dubois and Graham 2007 referring to Bloom 1998). While assertions as the just cited one might be true when looking at

productive vocabularies of sizes larger than 10, this assertion does not hold for the earliest productive and receptive words. Gopnik (1988) analysed the very first productive words of toddlers and found that social words are the consistently first type of words that are uttered at the onset of speech. She classified words from these early productive vocabularies as belonging into one of three categories: *social words*, *names*, and *cognitive-relational words*.

Social words are those "words that were consistently used to fulfill the same social function", with *social* indicating that they were always directed at other people. Gopnik furthermore noticed that these words are closely linked to illoctionary forces in the sense "that each word was always used to accomplish the same speech act, usually a directive speech act." Examples for these social words are *no*, used for refusal, *that* or *there* to point out objects, *bye* for leave-taking, *hello* or *hi* for greeting, *up* for requesting to be picked up, or *mama* as "all-purpose call for help or assistance".

Names are those words that occured in a variety of contexts and which could occur in a variety of of speech acts (unlike social words). Gopnik notes that "informal observation suggest that the affective context of these words was rather different from that of the social words", in that they often occured "when the child was not making eye-contact with another person and children seemed relatively unperturbed by a lack of response to these words."

Cognitive-relational words are then those words that "encoded cognitively significant concepts (other than objects or object categories), such as the concepts of success, failure, recurrence, disappearance and location" (comment in bracket added by author). Examples for these kind of words are *gone* to encode disappearance, *there*, *no*, and *uh-oh* to encode the success or failure of the child's plans.

For our purposes it might be worth emphasising that *no* which has two different functions here, *socially* as refusal, and *cognitively-relational* as comment on failure, as well as

gone, *cognitively-relational* as comment on disappearance, are all negation words.

2.2.2 Developmental Pragmatics

A field that certainly does not disregard the communicative functions of early child utterances, and in which researchers typically focus on a slightly earlier developmental period than the representationalists, is developmental pragmatics. According to Filipi (2009) the historical reason for the field's focus on the transition period between prelinguistic and linguistic communication were "the opposing positions of Chomsky and Piaget (Piatelli-Palmarini 1980)" on the relationship between communication and language and as to whether "Piaget's (1954) assertion that sensorimotor development provides the foundation for later developments in the child, including language, can be accepted" (Filipi 2009).

According to Filipi (2009) amongst the researchers that argue for functional continuity between the pre-linguistic and linguistic stage are Bates et al. (1977), Bates, Camaioni and Volterra (1979), Bruner (1983), Bullowa (1979), Carter (1975), Lock (1978), Werner and Kaplan (1963) and Zukow et al. (1982).
It is only against the background of this continuity assumption that terms such as *proto-conversation* (Bateson 1979, Bruner 1983), *protolanguage* (Halliday 1975), *proto-imperatives* (open-handed reaching), and *proto-declaratives* (attempts to elicit an adult's attention) (Bates et al. 1977, Bates, Benigni, Bretherton, Camaioni and Volterra 1979) have to be understood and make sense. These *proto* communicative acts are then seen as precursors to their later emerging, full-fledged and 'non-proto' linguistic counterparts.

Ryan (1974) is another researcher that emphasises the continuity between the two stages by arguing that the "(u)se and understanding of standard words develops at a time where the ability to communicate non-verbally is well estabished, in the sense of being

able to influence the behaviour of others, and of induldging in reciprocal interchanges of various kind" (p. 186).

With her criticism of quantitative-descriptive approaches and her emphasis on the interaction between mother and child Ryan might be one of the earliest adaptors of a conversation analytic approach to early child language development.

Developmental Perspectives on Speech Act Theory

Pragmatically leaning researchers of early child language development picked up very quickly the ideas of *(word) meaning as use*, that was developed by philosophers of language such as Wittgenstein, Austin, Searle, and Grice and asked the question, which the latter typically did not ask: How do children acquire the skills to express communicative intentions via conventional means, i.e. by the invocation of the appropriate illocutionary force? Another more fundamental question, which to the knowledge of the author, remains unanswered is then: How do communicative intentions come about in the first place and how do they relate to extra- or pre-communicative intentions, emotions, motivation, etc.? Bruner states this issue aptly:

> Grammarians usually take intent for granted but one does so at one's peril. ... But intent in communication is difficult to deal with for a variety of reasons, not the least demanding of which is the morass into which it leads when one tries to establish whether something was *really*, or *consciously* intended. Does a prelinguistic infant *consciously* intend to signal his displeasure or express his delight? To obviate such difficlties, it has become customary to speak of the *functions* that communication or language serve and to determine *how* they do so. This has the virtue, at least, of postponing ultimate questions about the 'reality' and 'consciousness' in the hope that they may become more

> manageable. (Bruner 1975a, p. 262)

Yet Bruner also asserts that the question when infants come to 'intent consciously' to communicate, i.e. when they move from "expressive utterances" ("early cries of discomfort and pleasure") to "stimulative" (producing reactions in others) and representational utterances, was not very useful as the criteria how one would judge when "that trip had been made" were not clear at all. The notions of *consciousness* and *intention* were opaque ones. Bruner (1975a) therefore suggests to focus on "how communicative functions are shaped and how they are fulfilled" in the mother-child interaction (p. 266).

Intentionality and early communicative intentions Dore (1975) supports the idea of conceiving of *communicative intentions* as linguistically expressed intentions that perform universal functions such as *asserting*, *denying*, *requesting*, or *ordering*. Yet he also contends that these "linguistically expressed intentions are not isomorphic with prelinguistic intentions and the former need not be derived from the latter" (Dore 1975, p. 37). Other proposed universals are *referring* and *predicating acts*, which are thought to emerge with the child's maturation and which are not deemed to be reducible to prior experience. This means, they are considered to be innate and, importantly, not derivable from pre-linguistic gestures for example (but cf. p. 62). He further contends that these non-grammatical entities become grammaticalised but are "cognitive prerequisites for the grammaticalization process" (Dore 1975, p. 38).

Filipi (2009), representing a more conversation analytical stance, sees communicative intentions as something that the child acquires due to the manner in which its caretakers interact with it: "Through a series of actions and counteractions, the infant learns that her actions or gestures have an effect. The parent interprets these gestures and responds to them as though they are words, as though the child has a social goal in mind. In

other words, through scaffolding and formatting, the parent is providing a conversational structure to the routines - "a prelinguistic communicative framework" (Bruner 1982, 1998) which is also referred to as scripts (Snow, Perlmann and Nathan 1987)" (pp. 8-9).

Intentional behaviour in its interpretation as goal-directed behaviour involving others can then be witnessed in a child "focusing on an object, making eye contact with the adult and/or gesturing while vocalising, trying an alternative behaviour if she fails in her goal and stopping these behaviours once the goal has been achieved, Bruner (1975b), ..." (Filipi 2009, p. 16).

According to Filipi (2009), Dore (1983) proposed as "(m)otivating force behind the intention to communicate ... the frustration of the child" while Garvey (1983) was taking a more positive stance by suspecting "the immediate satisfaction of having the means for producing an effect on others" as the source of motivation.

Intentionality Filipi (2009) identifies four issues in relation to the concept of *intentionality*, the first of which concerns its definition. As indicators of intentionality are taken to be:

1. the child's ability to control her actions intentionally,
2. the understanding of communicative intentions of others or the attribution of communicative intentions to others. This has been termed the ability to take an *intentional stance* (Carpenter et al. 1998).

According to Filipi supporters of this definition are amongst others Tomasello et al. (2007) and Liszkowski (2006). An obvious question, if we are to accept this definition, is then how we are to detect the presence of intentionality in the child. Filipi, summarising the results of research "from the perspective of the social context" of the previous few decades, lists the following four behaviours as indicative of the presence of intentionality (the given

references are only partial, see Filipi (2001) for further pointers):

- reaching and grasping for an object (Bruner 1973),
- gestural communication, especially pointing (Bates, Benigni, Bretherton, Camaioni and Volterra 1979, Filipi 2002, Tomasello et al. 2007),
- joint attention or reference (Bruner 1975b, Liszkowski 2006, Tomasello et al. 2007),
- attention getting practices, particular style of pointing or showing of object (Filipi 2001, 2002, Jones and Zimmerman 2003).

Intentional (or motivated) behaviour then might be equated with a theory of mind. This is done by firstly claiming that toddlers which display such behaviour were attributing beliefs and emotions to others (Carpenter et al. 1998, Golinkoff 1993: cited in Filipi 2009, p. 16), and, secondly, by equating the ability to perform such attributions with a theory of mind. (Golinkoff 1993: ibid.).

The other three issues surrounding intentionality as discussed by Filipi (2009) revolve around (1) the age of onset of intentionality, (2) the issue if the onset of intentionality is causal for the transition from pre-verbal communication to language, and (3) the question whether the empirical evidence that is typically presented when arguing for the presence of intentions "in" the child is actually sufficient to do so (Filipi 2009 referring to Francis 1979). We refer to the Filipi (2009, ch. 1) for a further discussion of these controversies.

In finishing our short discussion of *intentionality* we would like to highlight Filipi's (2009) distinction of the two different methodological approaches in laguage-related research that emphasise this notion. Filipi contrasts the so called socio-cognitive or socio-pragmatic approaches with the conversation analytical ones:

[1] The approach of Tomasello et al. (2007) and Liszkowski (2006) can be described as experimental, hypothesis-driven paradigm which is sometimes refered to as *social-cognitive* account of language development. Starting out with the assumption that a child recog-

nises itself as well as conversation partners as intentional agents, experiments under this approach are designed to show that this assumption holds. Problematic in this context is that it is unclear if insights gathered from experimental setups transfer to real life conversations (Schegloff 1992). Filipi (2009) considers it ironical "that these studies go to some length to show that pointing develops in the social context of shared understandings, yet we are simply presented with the reports of how the infants reacted. In other words, there is no relationship of the child or the adult's action to the interactional reality of the context in which the talk takes place" (p. 20). Thus, the actions under investigation in these 'usage-based theories of language' "are abstracted and removed from interaction and ultimately from a language in use paradigm" (ibid.).

[2] The approach of Filipi (2001, 2002), Jones and Zimmerman (2003), effectively an adaptation of Conversation Analysis to parent-child conversations, takes as its starting point the interaction as it naturally occurs. "A key methodological approach ... is to focus on the pursuit of their co-participating adult's attention both to themselves (the infants) and to the object for which mutual attention is sought" (ibid.). Further focus, in the spirit of the conversational analytic tenet that "it is in the next action" which is the "basic structural position by which participants' own interpretations and understandings of talk are displayed", "is on the pursuit of an appropriate, or at least adequate, response (Filipi 2007, Kidwell and Zimmerman 2007" (ibid.). The underlying assumption is that "cognitive processes implicated in the action of pointing and showing, including intention ..., are represented and emerge in interaction" (ibid.). In other words it is assumed that "constructs of cognition are "relevant to, and involved in, interaction in terms of their hearability in the interaction itself"" (ibid., p. 20: citing te Molder and Potter 2005, p. 24).

We would also like to highlight the slighty differing terminology between these two approaches. It appears that *intentionality* is the central notion of research conducted un-

der approach [1], whereas *intersubjectivity* appears to be the largely equivalent notion in research conducted under approach [2]. What is called "joint attentional frames" by Tomasello (2003) and Tomasello et al. (2007) translates to "socially shared knowledge or intersubjectivity Schegloff (1991)" within the conversation analytical terminology (Filipi 2009, p.19).

Operationalisation of communicative intentions Dore (1974a) contrasts *communicative intentions* (*CIs*) with *non-communicative intentions*, the *goals of an utterance*, and the *pragmatic purposes* of an utterance. He defines *communicative intention*, drawing on Grice's (1957) notion of utterer's meaning, as the "intention to induce in a listener the recognition of how the speaker wants his utterance to be taken". *Communicative intention* is then contrasted with Piaget and Cook's (1952) "sensorimotor intention", such as the intention to build a tower of blocks. In the latter case the child's intention and goal are basically identical, whereas communicative intentions involve other people.

This means that (communicative) intention, the inducement of the recognition of the child's plan in other people, is different from the *goal of an utterance*, which is the expected effect that this recognition will produce in the listener. If we compare these notions with those that are customary in speech act theory (cf. section 2.1.2), communicative intentions appear to be largely identical with the illocutionary level of a speech act because "the CI of an utterance is under the control of the speaker" (Dore 1974a, p. 6). The *goal of an utterance* then appears to be largely congruent with the perlocutinary level (or effect) of a speech act because "the goal of an utterance is under the control of the listener" (ibid.).

In this framework *pragmatic purposes* of an utterance are identified with rather indirect intentions and purposes that are not part of the normative function of a speech act. Asking somebody to do something in order to please a third person would be an example of such

a pragmatic purpose. Pragmatic purposes are thus related to indirect speech acts, which are as well beyond the scope of this literature review as they are beyond the scope of (Dore 1974a).

Typical names of communicative intentions of children are "questions, answers, labels, calls and protests". Dore goes on to report that he, based on a study with 3 year old children, identified the four "core" *CIs request*, *response*, *description*, and *statement*. *Descriptions* refer in this context to observable or verifiable aspects of the environment, whereas *statements* refer to unobservable "facts" such as intents, emotions, but also reasons, predictions, possession etc. Interesting for our purposes, none of the typical functions of negation is part of these "core" *CIs* unless one counts refusals as a form of negative request. Yet *protests* belong to the list of "non-core" *CIs*, a list which apparently expands as the child's development progresses, as opposed to the four types of "core" *CIs*, which according to Dore is likely to be fixed in number.

He further asserts that the non-core *CIs* "appear to be primitive versions of "performatives" ", whose felicity conditions are regularly violated by the children. An example of such violations of felicity conditions are 'broken' promises, where the child either does not follow up on what was promised, or already knows at the time of promising, that what it promises will come to be without further ado.

Primitive Speech Acts More or less in parallel to his treatment of communicative intentions, Dore also introduced in Dore (1974b) and Dore (1975) the notion of *primitive speech acts* (*PSAs*). He gives as reason for modifying the philosophical account of adult speech acts that "philosophers of language have not supplied an operational definition of *speech act*"(Dore 1974b, p. 344). His notion of *primitive speech acts* is then meant to help to answer the question how children acquire linguistic conventions. A second motivation

for postulating *PSAs* as central notion within the theory of early language development is Dore's hypothesis that children already had "sytematic knowledge about the pragmatics of their language before they acquire sentential structures" (ibid.) and that this knowledge was best captured by the notion of *primitive speech acts*.

This endeavour has to be seen on the historical background of the much discussed issue at that time as to whether or not children's one-word utterances "imply propositions" or "represent a sentence". This is a question which implies a primarily grammatical view of early language as opposed to a pragmatic one, as the sentence is given a pivotal role as meaning-bearing entity. The position, that single words during a child's one-word stage were 'really' propositions, lead subsequently to the notion of *holophrase* (Dore 1975). Dore argues against this predominantly grammatical perspective of early child language as "(t)he notion of sentence is the linguist's abstraction from adult utterances and may be totally inappropriate for describing early child speech. Yet the most studied problem of recent psycholinguistics has been the transition from the use of single words to syntax" (Dore 1975).

As a replacement for *sentence* or *proposition*, similar to the proposal of philosophers' of language in the context of 'adult language', Dore proposes the primitive speech act as central notion for the understanding of early child language. A *PSA* then is a combination of *rudimentary referring expression* such as "doggie", or "bye-bye", and a *primitive force* indicating device, typically an intonation pattern (Dore 1975). A *PSA* then consists of a one-word utterance or a single prosodic pattern, a consistent prosodic feature, and it communicates the child's intention (Dore 1974b). As opposed to holophrases *PSAs* are not elliptical adult speech in that they do "not contain a predicating expression" (Dore 1975, p. 32). "It expresses the child's intention with respect to a concept without having a propositional structure" (ibid.), and propositions are seen to emerge later with the de-

velopment of the child's grammatical capabilities. Dore further characterises *rudimentary referring expressions* as the child's ability to "linguistically represent a single concept".

The developmental trajectory from *PSA* to full-blown adult speech act is hypothesised to progress as follows. In the two-word-stage the rudimentary referring expression gets complemented by a predicating expression to form a *rudimentary proposition*. Additionally, the force component develops towards a higher degree of conventionality by being expressed "by elementary kinds of illocutionary force indicators" (Dore 1975, p. 34). During this stage the word order is still relatively random. After the two-word-stage the rudimentary proposition becomes grammaticalized towards a higher degree of conformity with the sentential structure of the target-language. Furthermore, the child's intentions are organised and are expressed as 'proper' illocutionary force, presumably within the modality component of the sentence (with *modality* being interrogative, imperative, negation). Intentions here then are "cognitive pragmatic structures, distinct from the grammatical categories that serve to express it" (Dore 1975, p. 36) and more akin to Piaget and Cook's (1952) notion of intention as "the deliberate pursuit of a goal by means of instrumental behaviours subordinated to that goal" (idid.).

In the study reported in (Dore 1974b) and which is based on 2 children in the one-word stage, Dore isolates eight distinct *PSAs* based on the following observational criteria: the child's utterance, its nonlinguistic behaviour (e.g. gestures, facial expressions), the adult's verbal and nonverbal response, and the situational context (e.g. salient objects, location, present people). The eight distinct *PSAs* that were distinguished in this manner are: *labelling*, *repeating*, *answering*, *requesting* (action or answer), *calling*, *greeting*, *protesting*, and *practicing*. Dore does not claim that this list is exhaustive. *Practising*, Dore's catch-all category for those utterances that could not be assigned to any other category, was the only category that was exclusively expressed in non-conventional forms, that is, with

linguistic features that are typically not used by adults. Also, adults typically do not practice word forms out of context. *Practicising* is therefore a speech act category that is typically not seen with adults. *Requesting* was done by both children conventionally and non-conventionally. *Protesting* was done by one child exclusively non-conventionally via extended screams with varying terminal contours, whereas the other child performed its *protests* exclusively conventionally during the experimental observations. Acts that were categorized as belonging to one of the remaining five categories were all performed conventionally by both children during the observation.

Dore further noticed considerable differences between the two children under observation where the girl engaged mainly in "representational" uses of language (*labeling*, *repeating*, and *practising* words) whereas the boy acted mainly via recognizable prosodic contours instead of using standardized words. The latter then used these idiosyncratic prosodic "words" mainly instrumentally, that is, to manipulate and influence other people by *calling* somebody, *protesting*, or *requesting* something. The mainly "representational" functions of the girl's early speech may be explained by the behaviour of her mother, who regularly set up routines for her daughter where she would pick up objects, label them, and encourage the child to imitate the label. This included animal-naming routines involving toy animals or animal pictures, as well as the naming of utensils and people.

More recently, more elaborate speech act taxonomies for child language have been proposed such as the Inventory of Communicative Acts (Abridged) (*INCA* and *INCA-A*) (Ninio et al. 1994). One of the questions that is necessarily encountered when trying to put speech act theory into practice, i.e. when trying to classify naturally occuring utterances from a conversation according to their illocutionary force, is the status of answers or other conversational turns whose meaning is dependent on a previous turn, essentially second

pair-parts of adjacency pairs. Searle (1969) remains silent on this issue, possibly due to his disregard for conversational phenomena (Searle et al. 1992). Ninio et al. (1994) solved this problem pragmatically by evidently giving these entities the status of proper speech acts: *INCA* also lists in all major categories of speech acts *responses* as members of the respective type.

Generally *INCA* is a hybrid taxonomy to code for communicative intent in that it is not only derived from speech act theory but also from Goffman's studies of face-to-face interaction (Goffman 1961, 1974) and from conversation analysis. Furthermore *INCA* codes communicative intent at two different levels, the utterance level (speech acts) and on the "level of the verbal interchange" which may consist of more than one rounds of talk "all of which share a unitary interactive function" (Ninio et al. 1994). Moreover, the authors of *INCA* tested the 'ecological validity' of the taxonomy by using mothers as informants about, and interpreters of, the social reality created in the dyadic interaction.

In terms of negation we find "refuse to carry out act requested" as separate speech act type in the category of directives in *INCA-A*.

Developmental Perspectives on Conversation

Conversation analytical approaches are generally very critical about theory-driven research methods, and thus do not only critisise what may be called "positivist approaches", i.e. approaches that focus by and large on truth-functional propositions, but they may even critisise pragmatic approaches for similar reasons: "Positivist approaches that rely exclusively on count data can ... be criticised (as) they do not tell us what those features are doing and how they contribute to the overall organisation of talk" (Filipi 2009). Yet, according to Filipi, these, together with speech act theoretical approaches, have dominated the studies of infant talk . The problem even with pragmatic approaches for conversation

analysts is that also with the latter it is "linguistics that provides the starting point ... rather than the talk *per se* viewed as sequentially organized social actions" (Filipi 2009, p. 53: refering to Sharrock and Anderson 1987). In other words these approaches are theory-driven and thus use "preconveived notions of talk and premature categorising based on intuitive judgement" (ibid.).

Conversely conversation analysis is data-driven and "there is no research question or hypothesis at the outset" such that "nothing can be dismissed *a priori*" (Filipi 2009). This stance has been described by Levinson (1983) as "a strict and parsimonious structuralism and a theoretical asceticism" (p. 295).

Intent Interpretations Joanna Ryan, who appears to be one of the early adopters of a conversation-analytical position within child-language research, emphasises the adults' role in and impact on language acquisition by asserting that "(e)arly language development thus appears to take place in a context that provides a child with frequent interpretations of his utterances" (Ryan 1974). She observed that mothers frequently repeated or extended the child's utterances and also manipulate the non-linguistic context in order to understand what they conceive of as attempts at speech. Several features of the child's behaviour are observed to contribute to adults crediting the child with 'trying to say something', which they also use to make sense of the utterances. The behavioural features presented by Ryan are intonation patterns, which within Dore's account featured as primitive IFIDS, and which are "variously interpreted as insistence, protest, pleasure, request, etc." (idid.) based on adult intonation patterns, accompaniments of utterances such as pointing, searching, refusing but also other features of the situational context such as the presence of objects or people that have or had some relevance to the child. Important for our purposes is the observation that these linguistic interpretations do not only oc-

cur as extensions of single words or some non-conventional but stable phonetic pattern, but also as interpretations of bodily behaviour. According to Pea (1980) "(p)arents also frequently interpret these behaviors (physical means of rejection, our comment) as expressive of negation and expand them with lexical negatives: "no, no, don't want it."" More recently, Filipi (2009) provided another, potentially complementary explanation for these behaviours which play a central role in our hypothesis 1 (cf. section 1.1): parent(s) "address the infant as a conversational partner from birth and continue(s) to do so through early infancy. As noted, the parent treats the infant's gestures, eye contact and vocalisations as meaningful and she will respond to them verbally. *When the child does not produce the required conversational behaviour, the parent supplies it*" (Filipi 2009, p. 23, emphasis added). If this is the case, and we will see clear evidence of such behaviour in our human-robot dyads within the experiments, the caretaker's activity might go well beyond the mere interpretation and completion of the child's incomplete utterances. Caretakers and, in our case, participants, may provide the entire conversationally required turn, if the child (or robot) does not appear to fulfill its obligation.

Ryan cautions us against the practice to "conflate the means by which an adult interprets a child's utterance with the devices it is assumed a child uses to express herself", that is "to assume a child uses the same devices to convey her meaning as an adult uses in interpreting the utterance" (Ryan 1974). This problem is fundamental to the application of pragmatic theories, speech act or other theories, to mother-child conversations. This fundamental limitation also touches upon the basic assumption of ethnomethodology and conversation analysis that "unique adequacy can usually be assumed" for "the subject of research is something that most persons participate in regularly, like ordinary talk" (Rawls 2003). In other words we, as researchers, may not be able to access the *member's methods* of the child by means of "naturalistic observation grounded in a deep familiarity with and,

preferably, bona fide competence in the discipline under scrutiny" as our familiarity with ordinary talk gives us access to the mothers' member methods but not necessarily to those of the child (Ryan 1974). Yet researchers such as Trevarthen evidently trust our capacity as conversing human beings to be able to interpret a toddler's behaviour in a meaningful way, in his case by means of a (conversational) micro-description of mother-child interactions (cf. Trevarthen and Aitken 2001). Conversely it may be argued that the interpretation of other reseachers of a toddler's single-word utterances as being just words or as being proto-utterances was theoretically motivated and cannot be backed up experimentally neither. We will later follow Trevarthen's lead by using a ethnomethodological approach in our analyses of the pragmatic level (cf. section 5.3).

In the following we will give a rough sketch of pre-linguistic adult-child communication based on (Filipi 2009).

Early Communication - Gaze As first step into a social world as conversation partners, neonates have been observed to orient towards a talking adult within the first week.

Gaze is generally the first means of infants to signal to an adult that he shall attend to an object which is accomplished via *gaze fixation*. In this function it is thought to be the first developmental step for communicating attention and as such a precursor to, first, gestural requests and subsequently verbal requests (p. 2). This ability subsequently changes the mother's behaviour in that she starts to communicate with her child. She "fits in" with the infant's behaviour and responds to her, while ceasing to respond when the baby looks away (Stern 1974, Brazelton et al. 1974, Fogel 1977, Kaye 1979: ibid., p. 3). The roles of the 'gazers' are initially asymmetric: parents were observed to watch the child much more than vice versa. Parents have been described as acting as if the child was behaving intentionally (Bruner 1975b, Collis 1977, Harding 1983, Trevarthen 1979: ibid., p. 3).

Apart from gaze, a sharing of rhythm both in talk and body movement between parent and infant have been observed (Holmlund 1995: ibid., p. 3) and have been described as the "groundwork for the future development of turn-taking" (Bateson 1979: quoted in Filipi 2009, p. 3). Thus gaze may be seen as the "infant's way of starting to do "interaction" or of setting the stage for talk ... The parent's role in this is of course crucial. It is largely through her behaviour that the parent systematically "teaches" the child the importance of *gaze in turn-taking*, and that the basic sequential organisation of talk involving *adjacency pairs* ... is beginning to set" (Filipi 2009, p. 3, emphasis added). Filipi contends that the child beyond acquiring rules for turn-taking actually learns that gaze is a form of social action. Yet, even at this very early stage, universality may not be given: "sustained gazing and the importance of maintaining eye contact is very much a feature of parent and child interaction in middle class Western families" (p. 3). Amongst the Kaluli, a community that lives in the Southern Highlands province of Papua New Guinea, gaze appears to play a far less pivotal role in mother-child interaction than in Western families (Schieffelin and Ochs 1983).

Interestingly face-to-face interactions occur significantly less after the first six months as the child learns to follow the direction of the parent's attention. The latter is thought to be very important for the later lexical development, as joint attention and the ability to establish a joint focus are claimed to be crucial in this context by authors such as Tomasello (2003). By 10 to 11 months toddlers then acquire the skill to follow the adult's eye direction, before that stage they are following the adults' head movements (p. 4).

Early Communication - Gestures Proponents of the continuity assumption, i.e. the assumption that gestural and linguistic communication form a unified system, claim that "what starts off as a spontaneous gestural action in the early stages of infancy, becomes

intentional communication at around nine to ten months" (Bates, Benigni, Bretherton, Camaioni and Volterra 1979, Caselli 1990: ibid., p.8). According to these, infants are unaware of the "conventional purpose" of their gestures up to that stage (Carpenter et al. 1998, Liszkowski 2006, Liszkowski et al. 2007, Tomasello et al. 2007: ibid., p. 8). Deictic gestures, mainly concrete pointing, occur first, with "intentional pointing" starting at around 9 months, followed by conventional and iconic gestures such as head shakes or waving (Filipi 2009, p. 10).

Recent studies indicate that children's first utterances are crossmodal (Gullberg et al. (2008): ibid., p. 11), thus, gestural communication does not stop with the onset of speech. Some authors have linked the early comprehension of words to action gestures (Caselli et al. 1995, Fenson et al. 1994: ibid.). Pointing and reaching gestures in particular are thought to be of great importance to early language development because of their communicative function (Leavens and Hopkins 1999: ibid.). Some researchers even attribute to gestures the same linguistic status as to the first words, and Bates et al. (1983) go on to equate gestures with nouns (Filipi 2009, p. 12).

Open-handed reaching, which we will employ in our experiments, was described by Bates et al. (1977) as a 'proto-imperative', through which the child is making her earliest requests. The parents' response to this gesture then may give it the appropriate information on how to influence others which then might eventually lead to the development of the imperative (Filipi 2009, Bruner 1983).

Development of Turn-taking Possibly the earliest precursor of turn-taking as observed in talk-in-interaction has been described in the context of infant feeding: there an "alternation between sucking, pausing and looking at the mother who then engages in talk" (Filipi 2009, p. 24) has been oberved and authors such as Kaye (1977) consider this sequential

organisation a conversational prototype (ibid.).

Yet, if one is willing to accept this position, the sequential organization is no "neat alternation of speaker acts" at that stage. Overlaps of vocalizations occur frequently and have been given the name "vocal clashing" (Ely and Gleason 1995: ibid., p. 25). Vocal clashing peaks at 7 to 13 weeks (Locke 1993) whereupon turn-taking becomes more orderly such that at the end of 12-18 weeks infants adapt in that they abstain more frequently from vocalising during productions by their mother (Symons and Moran 1987): ibid., p.25). This is then the time period when terms such as *proto-conversations* (Bateson 1979: ibid.) and *pseudo-dialogues* (Schaffer 1979, 1984: ibid.) are used to refer to the interaction between mother and child. Some authors claim that it is the mother's conversational work, i.e. her adaptations to the timing of the infant seeking to minimise overlap and her treatment of the child as active conversational partner, that creates the impression of a real conversation taking place (Filipi 2009 refering to Hayes 1984, Schaffer 1979, 1984). Other authors such as Murray and Trevarthen (1986) criticise these accounts as withholding the active role of the infant and therefore denying the reciprocity in this synchronization process (Filipi 2009).

Most responses in parent-child conversations are produced within one second (Beebe and Stern 1977, Stella-Prorok 1983: ibid., p. 25) with gaps after maternal utterances typically being longer (Snow 1977: ibid.). This is an interesting observation as the one-second threshold appears to be an important threshold in adult conversations (cf. section 2.1.3).

In the context of this thesis possibly the most important observation is that the child's turn-taking skills are already well established when first words are uttered: "... when the child reaches the stage where she can utter her first word, she is already capable of sustaining long periods of well-timed turns at talk" (Filipi 2009 with reference to Kaye and

Charney 1980, 1981).

Adjacency Pairs In section 2.1.3 we have already emphasised the importance of adjacency pairs for the sequential structure of adult talk. On this background it is an important observation that question and answer sequences are pervasive in early turn-taking in parent-child interaction (Ervin-Tripp 1977, Snow 1977: refered to in Filipi 2009, p.26). In this context the main focus in the literature, most of which was published in the 1970s, is on parent's questions. It was observed that "asking questions is more characteristic of a parent's talk to her child" than vice versa (Filipi 2009, p. 27: refering to Keenan and Schieffelin 1976). Furthermore "(t)he high frequency of questions is maintained in the parent's interactions with the child throughout her early life, from the age of three months to three years" (Filipi 2009, p. 27 with reference to Johnson 1982). The hypothesized reason for the high frequency of questions in parent's child-directed talk is to generate and sustain a conversation as the production of first pair-parts create a slot for the child to respond (ibid., p. 28: refering to Erwin-Tripp and Miller 1977, French and Pak 1991. Indeed parental questions "have been found to be more successful than comments as elicitations (Yoder et al. 1994)" in that "they are more likely to receive a response (Foster 1979)" (ibid., p. 27).

What can be thought of as cognitive complexity of question-answer pairs changes during the child's development from cognitively easier tag questions to questions that demand specific responses, requests for action and test questions. Furthermore the length of parental repetition sequences of questions decreases and children "are increasingly able to terminate them with a relevant response (Snow 1978, Filipi 2001)" (ibid., p.28). An interesting observation in this context is that "answering questions is thought to be the first discourse bound obligation to which the child is sensitive (Erwin-Tripp and Miller 1977)" (ibid., p.

29). This is not to say that children always fulfill this obligation satisfactorily, "... they might simply imitate the adult by repeating the prior utterance (Baker and Nelson 1984) or randomly choose between *yes* or *no* despite the fact that the answer may be incorrect (Tanz 1987)" (ibid.).

2.3 Developmental Robotics

Developmental robotics is a relatively new research field within robotic research that is inspired and guided by insights from cognitive and developmental sciences, where the two developmental sciences under consideration are developmental psychology and developmental neuroscience. Lungarella et al. (2003) describe its methodology as two-pronged in that it instantiates models that originate from the developmental sciences on one hand, and develops better robotic systems by exploiting insights from these sciences on the other hand. Robots are thought to be valid tools to investigate embodied models of development in the expectation that psychologist "may gain considerable insights from trying to embed a particular model in robots" (ibid.). Moreover, the field is seen not only as very similar to but rather as an extension of epigenetic robotics, in that it endorses an biomimetic approach by addressing biological questions via the construction of physical models of the animal in question (ibid.). One may add to this view, that 'the animal' might be human, and that in such cases 'biology' extends to include psychology. The 'add-on' of developmental robotics as compared to epigenetic robotics is then the additional consideration of research questions pertaining to the acquisition of motor or other skills, i.e. an extension of scope to include ontogenesis.

Robots that are employed within this methodology are "'cognitive' or 'synthetic' research tools ... to study and model the emergence and development of cognition and

action" (ibid.). As extension of epigenetic robotics the field then inherits its emphasis on the importance of embodiment. Thus, most researchers of this field appear to reject the mind-as-computer metaphor of traditional cognitive science, which demotes the body to a mere 'slave' of a disembodied mind, and where the body serves either solely as channel for sensorial inputs or as a mere output device or 'executor' of what the independent mind had 'decided' (cf. Varela et al. 1991). This means that epigenetic and developmental roboticists generally oppose the view that all cognitively interesting things happen detached from all bodily matters, but, conversely, see the body as a prerequisite and active element in the generation of intelligent behaviour. The traditionally strong separation in robotics between cognitive structures (symbols, representations), software (attention, decision making, reasoning), and hardware is viewed as rather unfortunate (ibid.). Lungarella et al. (2003) further attest the cognitive approach a denial of the importance of ontogentic development and the (linguistic) nativist position is seen as one outbirth of the cognitivist stance.

The behavioural system that constitutes a central element within this thesis (cf. section 3), thus may be viewed as a developmental robotic system, as the hypotheses which are tested are developmental ones, pertaining to the acquisition of negation, but also because we followed a particular tenet when designing the system, that Lungarella et al. (2003) consider as basic within the developmental synthetic methodology: "the designer should not engineer 'intelligence' into the artificial system ...; instead, he or she should try to endow the system with an approriate set of basic mechanisms for the system to develop, learn and behave in a way that appears intelligent to an external observer" (p. 179).

2.4 Robots and (Human) Language

Possibly *the* central idea within the community of 'robotic language learners', by which most if not all embodied approaches distinguish themselves from non-embodied, computational approaches to language understanding and learning, is that they attempt to solve the so called "symbol grounding problem" (Harnad 1990), depicted in figure 2.1. If one accepts

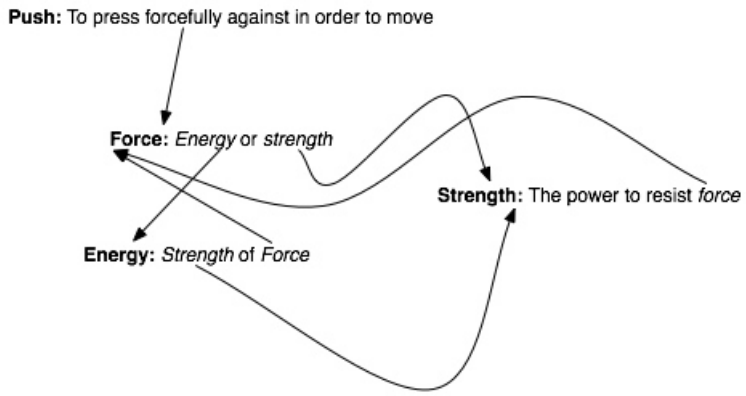


Figure 2.1: ***Symbol grounding problem****: One cannot learn the meaning of words from a dictionary alone due to the depicted "dictionary go-round" (aka circular definition of lexical items). Without symbols being* grounded *in something outside of the dictionary, the learner will necessarily end up in such a go-round. The depiction is a (non-faithful) replication from a depiction given in Roy (2005), which is based on Webster's Dictionary.*

this problem as central to the acquisition and understanding of language by an artificial agent, all existing approaches to language acquisition, learning, understanding, or recognition fall into one of two categories: those who try to tackle the problem, and those who don't. The vast majority of modern computational approaches to speech and language processing such as natural language processing (NLP) or automatic speech recognition (ASR) (Jurafsky and Martin 2000) fall into the latter category. An alternative approach to avoid the grounding of symbols in perceptual or 'bodily' data are so called *amodal symbol systems*, with a popular exponent being latent semantic analysis (Landauer and Dumais

1997). Due to the lack of space, we cannot discuss this approach any further, but refer the reader to (Barsalou 1999) for a critical discussion.

It has to be emphasised that most of the latter approaches do not attempt to make 'their' machines understand human language in more than the most superficial ways. A search engine, for example, does not necessarily need to 'make sense' out of the data that it is processing. There, it might be sufficient to generate a list of results based on a few catchwords entered by the user and based on the behaviour of other users that 'asked the same question', i.e. entered the same query. Similarly, a speech interface for cars, if the car is sufficiently modern, would theoretically have access to more 'external world' sensors than most modern robots. Yet it does not necessarily need to understand the user input in any 'deep' way, in order to do its job. What this interface needs to do is to call the correct function, based on the user input, and where the correct speech-to-command mapping is predefined by the designer. There is no real need for the car to have 'acquired' these mappings, it is sufficient if the designers of the interface had enough foresight to hard-wire the vast majority of utterances that a human would possibly utter when 'speaking to' the car. The latter could presumably be lifted from an extensive user study. Only a few people would presumably try to converse with their Mercedes about the latest football game, yet those who tried would be utterly disappointed by the outcome of the attempt. Furthermore, users will most probably be more than happy to adapt their way of speaking to a certain degree, in order to 'be understood' by their vehicle, i.e. they will stop their attempts to converse about football. The same holds true for modern speech interfaces for mobile phones. With enough data of human speech behaviour available, and this data is growing by the minute, there is no real practical need for command interfaces to 'really understand' a single word of what the user says.

Yet, this is precisely what, at least in the author's opinion, symbol grounding is all

about: for a machine to make sense of, first words, and then, hopefully, longer utterances, by virtue of linking these words with its own embodiment. It is important to emphasise the active construction in the previous sentence: "linking" instead of "being linked". Because one could easily argue that a speech interface for automobiles that does map words or utterances to control functions does indeed have some form of symbol-to-percept or symbol-to-action grounding. One could easily imagine a function that evaluates the status of the rain sensor on the car window in order give a meaningful answer to the question "Is it raining outside?"[10]. This utterance then would be grounded in the best sense of the word. Yet, the link between the sensory-function and the word or utterance was in this case established by the designer, not by the car itself.

We should therefore be more precise with regards to what those roboticists that attempt to "ground symbols" are actually attempting to do. The issue that is being tackled is how to enable the machine to learn or acquire the symbol-to-percept or symbol-to-function mappings by itself. To use the car example: How can the Mercedes figure out all by itself that "rain" somehow relates to its water sensors as opposed to the sensor that measures the engine temperature. The major challenge, thus, is not the grounding in the sense of the resulting link itself, but rather how to have the machine establish this link in an automatic fashion.

This then is the challenge that is tackled by most researchers in the field of 'robotic language acquisition'. Yet, another much less recognised and tackled problem is the circumstance that utterances are not just strings of words. As alluded to in previous sections, utterances also have a communicative function. This means that questions are not the same as assertions. "Do you know what time it is?" can be a proper question, yet it also can be a request to be left alone. And this function evidently does not hinge on words or

[10]It is another question if this actually makes any practical sense, as looking out of the windows would probably answer this question more quickly than asking the car.

strings of words. Do then communicative functions have to be grounded as well? And if yes, how? What are these communicative functions then mapped or linked to? And how many communicative functions are there anyway? To the knowledge of the author, there have been hardly any attempts in the robotic community to tackle this issue. Yet this is precisely the problem, which we face, when trying to 'ground' negation.

2.4.1 Previous approaches to symbol grounding

In this subsection we will quickly summarize the previous attempts to symbol grounding via conversation or dialogue and work out important differences between those approaches. This subsection is not meant to be a comprehensive overview due to time and space constraints. It should be noted, that not all symbol grounding is done via conversational systems. Evolutionary linguistics is a neighbouring and methodologically somewhat overlapping field, where this notion is equally central, yet the perspective there is an evolutionary one, phylogenetic as opposed to ontogenetic, and typically involves multi-agent models, often simulated (cf. Cangelosi 1999, 2001, Lyon et al. 2007, Steels 2003, 2005).

We further sub-divide this subsection in terms of the nature of the conversation partners, i.e. in terms of who is conversing with whom, as this has important implications on the complexity of the grounding problem. Notice that the section on "Designer-Robot-Interaction" comprises experiments in which the human interlocutor has to restrict his or her speech in pre-defined ways in order to render the symbol-grounding successful. Thus the section title is somewhat of a misnomer and was chosen for reasons of brevity.

Robot-Robot Interaction

Some of the embodied frameworks such as those described by Steels (2003) enable artificial agents to invent their own vocabulary and simple forms of grammar like word order for the

purpose of communicating with each other in a language constructed by and understandable to the robotic participants of the dialogue (Steels 1998). The latter approach came to be known as *evolutionary linguistics* and mainly focuses on the development of vocabularies over time by groups of agents, and potentially, simple grammatical constructions by the agents (Steels 2005). As the research effort in the involved approaches appear to be mainly directed towards the understanding of the evolutionary dynamics of the language games of the agents, the symbol-grounding problem is often simplified by limiting the number of potentially meaningful sensory channels, often the visual channel (e.g. Baillie and Ganascia 2000). Most often these systems are limited in their scope of acquisition to object labels and descriptions of physical actions or events such as *move*, *push*, *pull* etc. (Steels and Baillie 2003). Often the underlying conceptual layer, i.e. the semantic equivalents to which event 'labels' such as *push*, *pull*, etc. are linked is hand-crafted by the designer and, in turn, linked to some event logic (e.g. Steels 2003). This means that what is typically learned, is the word-to-concept mapping, whereas the concepts are already existent, static, and can typically not be influenced by the linguistic level.

A popular 'template' for these logical type of grounding systems appears to be the one developed by Siskind (2001), variants of which were for example adopted and/or developed by Dominey and Boucher (2005) and Steels and Baillie (2003).

Siskind's (2001) grounding mechanism consists of several levels:

(1) A *segmentation-and-tracking* component that isolates and tracks coloured objects on the camera images.

(2) A *model-reconstruction* component that produces a so called force-dynamic model for the objects as identified by the segmentation-and-tracking component. The latter determines which of a number of visual primitives apply to the isolated objects. Visual primitives in Siskind's system are *grounded*, i.e. physically supported by an unseen mechanism other

than the known objects, or *attached*, which indicates that the object is attached to another object in some way.

(3) An *event-classification* component that determines over which time intervals certain primitive events hold based on the force-dynamic model. Primitive event types are then events such as SUPPORTED(x), SUPPORTS(x,y), CONTACTS(x,y), and ATTACHED(x,y) which form the basis for an inference mechanism which can subsequently determine higher-level events such as PICKUP(x,y,z), or MOVE(w,x,y,z).

The high-level events can then subsequently serve as semantic basis against which object and event labels may be linked.

Obviously, grounding systems of this kind are limited to the grounding of those physical and visible actions and events which are decomposable into the respective visual primitives on one hand, and for which inference rules have been designed by the human constructor to 'perform' this decomposition on the other. The grounded utterances are typically utterances such as "the red block pushed the blue ball" and the like.

Designer-Robot Interaction

(Steels and Kaplan 2002) present a so-called 'social learning' mechanism where the Sony robot AIBO engages with a mediator in a classification game, effectively a game that aims to establish an object-word association - a game not dissimilar to the one that participants and robot engage in within this thesis (cf. section 4). The 'social' component of the game, i.e. the feedback of the mediator consists of encouraging and correcting words such as "good", "yes", or "no", which are pre-determined and not learned. Furthermore the mediator obviously has to be trained in order to only utter trigger words and utterances which activate scripts that, in turn, drive the interaction. On the plus-side the learning evidently happens in real-time, which is made possible by the pre-arranged scripts, an

automatic speech recognition which is trained for the small corpus of words and utterances, that the mediator is allowed to utter, and a mediator that sticks to these limitations.

(Dominey and Boucher 2005) focus on the acquisition of more complex sentence-to-meaning mappings by using a camera-setup coupled with a speech recognition system. Content words and their order are extracted from (complete) sentences via the identification of pre-defined function words. The extracted words and their order is subsequently mapped to a predicate-argument structure that is derived from a physical scene via an event analysis based on perceptual primitives akin to the one developed by Siskind (2001). These mappings are then stored in an associative memory with the configuration of function words as index to access the stored mapping. The authors subsequently tested the ability of the system to generalize across participants, i.e. to identify correct noun-agent, noun-object, verb-action mappings etc. based on unseen sentences. This system is thus very much motivated by construction grammar approaches such as (Tomasello 2003), yet it also assumes that the sentences are complete and that they exclusively refer to the physical scene at hand. Systems of this kind are, despite being 'constructionist', in the grammatical sense of the word (Tomasello 2003), prime examples of representationalist approaches to language acquistion, i.e. they assume that every utterance is a complete sentence and that the sentence 'represents' or corresponds to a physical scene. Albeit the authors tested their system with so-called naïve participants, these participants were constrained to one type of speech act: describe a given physical scene. Any other kind of linguistic behaviour has to be either captured by the foresight of the designer, or the human has to be forced to speak in certain system-compliant ways. We therefore do not count this system amongst those that involve unconstrained human-robot interaction but rather amongst those where the speech is somewhat designed to fit the learning problem.

Unconstrained Human-Robot-Interaction

The experimental approach presented in this section forms the basis of the work described within this thesis and will therefore be discussed in greater detail than the other robotic approaches sketched above. The major difference between this and the other presented robotic approaches for language-centered human-robot interaction is the circumstance that here naïve participants are the ones interacting with the robot, and that these participants can speak with the robot in whatever way they deem appropriate. Furthermore no post-experimental filtering of the collected speech is undertaken to remove 'unfitting' utterances or the like to improve the learning algorithm. This 'lack of tampering' with the data and the lack of scripted dialogues, which appear to be common amongst the systems in our *designer-robot interaction* category, renders approaches of this kind most similar to the learning problem that a small child faces and more informative in terms of the characterisation of the actual problem that children face. Naturally this large degree of 'interactive freedom' also renders the learning problem comparatively hard, as there is no guarantee that the utterances really fit the experimental scenario.

Saunders' system Saunders et al. (2009) describe a two-pronged approach to robotic language acquisition with the first 'prong' focussing on pre-word simulated babbling and the second 'prong' focussing on the developmentally later acquisition of first words in an "interactional environment with shared 'intentional' referencing" (ibid.). We will focus here on the latter avenue of research and refer the reader to (Lyon et al. 2012) for a discussion of the simulated babbling approach. The latter approach was in terms of the employed learning method set out as one of establishing a statistical association between speech and the robot's actions, its visual, proprioceptive, and auditory perceptions. As the authors outline, a 'brute force' statistical mapping that operates indifferently on all available speech

and perception data is most likely to fail, as observations of mother-child conversations do not support the view that the language teacher always utters the 'right' words at 'right' time to establish the correct link between the extralinguistic object and the intralinguistic word. For this reason the learning method is biased via the establishment of a shared context and shared intentional 'intent' between robot and human teacher.

Experiment 1: Acquisition of noun-like object labels For the first experiment the social learning architecture *ROSSUM* (Saunders et al. 2006, 2007) and the humanoid Kaspar2 (Blow et al. 2006, Dautenhahn et al. 2009) are employed. The robot's behaviour is driven by novelty, i.e. it searches for, fixates on, and tracks novel objects, and smiles as soon as one is found. Yet the robot also becomes bored with a given object if it is presented long enough (approx. 20 sec.) and subsequently moves its head semi-randomly until a new object or the human's face enters its visual field. This behaviour was designed in order to provoke participants to present new objects to the robot. The robot is further driven by the 'urge' to share the same attention space as the human participant and its visual focus thus is made to correspond roughly with the target of the human gaze.

In Saunders et al. (2010) the authors report about the outcome of an experiment based on said architecture and robot. The experiment was initially set up to make use of an automatic speech recognition (*ASR*). Yet the accuracy of the word recognition of these systems was found to be insufficient for the purpose of lexical grounding, and an alternative semi-automatic method for speech processing was developed (see next paragraph). The experiment consisted of 8 participants, each of which completed four training sessions for the *ASR*, followed by five interaction sessions with the robot of approximately 2 minutes each. Participants were told to teach the robot the names of the given objects and to to treat the robot as a 1-2 year old child. For the purposes of symbol grounding an 8-

dimensional sensorimotor (*sm*) vector was used, consisting of the object id, a binary value for face detection, head pan, tilt, and roll and 3 dimensions pertaining to the location of the object.

Speech processing, word extraction, grounding - Method 1 The speech processing method developed by and employed in Saunders et al. (2009) was adopted by us and we therefore refer to section 3.6 of the architecture for a description. However, as opposed to the word extraction method employed within the work of this thesis, Saunders et al. (2009) employed a different heuristic to determine the most salient word of an utterance. Under the latter heuristic a word is considered the most salient word of an utterance if (a) the duration of its pronunciation is above the average word duration within the utterance, and (b) if the word is at the utterance-final position. For each utterance at most one salient word is extracted. The grounding of salient words, i.e. the 'attachment' of sensorimotor data to the extracted salient words was, again, adopted by us, and is described in section 3.7. Naturally Saunders et al. (2009) system differs from ours in terms of the kind and dimensionality of the sensorimotor data, yet the grounding process is identical.

Run-time 'languaging' The method by which words are selected from the robot's lexicon, i.e. the set of grounded salient words, was largely adopted by us and is therefore described in section 3.4. However the two systems differ in the following four aspects: (1) Saunders et al.'s (2009) system uses information gain to 'weigh' the sensorimotor dimensions for distance calculation within the *kNN* algorithm, (2) Saunders et al. use the 1-norm Manhattan distance as similarity metric, (3) the 'speaking threshold' is not adapted to the robot's motivational state as their system has none, and (4) no differential lexicon is employed (cf. section 3.4).

Experiment 1: Research questions and answers Four questions were sought to be answered by Saunders et al. from the experiment. The first two questions, pertaining to potential adaptations of the participants speech, were [*1*] if participants in the experiment would adapt their speech when interacting with the robot in a similar way to adaptations observed in CDS, and [*2*] whether participants would change their speech style with the progression of the experiment during which typically a change of the robot's speech occurs as its learning progresses. The last two questions pertain to the robot's learning progress and were [*3*] whether the robot's 'linguistic classification' would improve as the experiment progresses, and [*4*] whether the robot would 'hone in' on those dimensions of its sensori-motor data which we, as knowledgable observers, know to be relevant for the classification of the present objects.

With regards to question [*1*], Saunders et. al. found, that female participants lowered their speech rate across sessions, whereas male participants did not. Moreover it was found that participants placed the relevant words, that is the object-related nouns, at the end of the utterances, and pronounced them with higher-than-average duration in, on average, 80% of times.

The authors were seeking to answer question [*2*] with an analysis of the number of added and dropped words between sessions. This analysis indicated a considerable amount of repetition. Furthermore, the largest drop in the number of words was witnessed between the second and third session, i.e. after participants had heard the robot speak for the first time. The interpretation of the authors with regards to this finding was that participants were adapting to what was most probably perceived as a limited understanding of the robot.

With regards to the improvement of the robot's 'linguistic classification', the authors found that such an improvement did indeed happen between the first and the fourth

session, such that the percentage of 'correct matches' were ranging between 53 and 76% after this fourth session. Yet this correct match rate dropped from the forth to the fifth session uniformly across participants. This was explained by the following adaptation: upon the mainly correct production of object labels on part of the robot by the 4th session participants often stopped to teach it the names or object labels and proceeded to give it encouraging feedback such as "well done". This positive feedback then, eventually, leads to the entering of 'object-foreign' words into the robot's lexica, a process which subsequently dilutes the association between the object-id dimension of the *sm* data and the object labels.

In order to answer question [*4*], an analysis of information gains associated with the various *sm* dimensions was performed. This analysis showed that for all but 2 participants the information gain associated with the *sm*-dimension that contained the object id was indeed higher than the ones associated with any of the other dimensions. This then indicates that this dimension is the most meaningful dimension with respect to the words in the lexica.

Experiment 2: Moving towards the two-word stage Saunders et al. (2012) describe the attempt to move the robot's language production from the one- to the two-word stage. This target necessitated a slight change of the experimental scenario such that the previously black-and-white and equally large boxes were replaced by coloured boxes with different sized signs on them. This change was made in order to enable participants to talk not only about the objects as wholes but also about their attributes, here colour and size. The participants were further told to do precisely that: teach the robot about colours and sizes on top of the 'kind' of the object. Another, technical, change in the experimental setup is the exchange of the robot. Kaspar2 was replaced by the humanoid iCub, the same

humanoid that was also employed in the experiments reported about in this thesis. The robot's behaviour was slightly modified as well: deictic gestures (pointing) and non-deictic arm movements towards the objects were introduced. Furthermore happy and sad facial expressions were employed and small random head movements to engage the participants. The sensorimotor data employed was reduced to three dimensions, corresponding to object id, colour id, and shape id.

Speech processing, word extraction, grounding - Method 2 While the speech processing in this 'advanced' system still is done in the same manner as has been described for the first experiment above, the saliency detection was modified in order to extract more than one salient word from each utterance. This new method employs prosodic features and preliminary experiments on this method are reported in (Saunders, Lehmann, Sato and Nehaniv 2011). In order to obtain a yardstick to determine which words can be considered prosodically salient, the product of the normalised values for *maximum fundamental frequency* (*f0*), *duration*, and *maximum energy* was calculated for each word. A word was subsequently considered salient if this value was larger than the average value of this product within the utterance to which this word belonged. Thus, as opposed to the 'old' method, the 'end of utterance' criterion was dropped and replaced by the two mentioned prosodic measures, while the word duration was kept as determining factor for salience.

The grounding is performed in the same manner as before. Yet, there is now potentially more than one salient word per utterance. If this is the case, the salient and grounded words are split into two sets or memory tables. The first table then contains all but the last of the salient words of any utterance, whereas the second set contains the last salient word of each utterance.

Run-time 'languaging' - Method 2 The method by which words are retrieved from the memory tables is essentially the same as before, yet there are now two tables against which the incoming *sm* data is matched simultaneously. Thus, there is now the possibility that two words, one from each table, might reach the 'expression threshold' at approximately the same time. In this case both words are produced by the robot in the order *<word from table 1>*, *<word from table 2>*.

Experiment 2: Research questions and answers Saunders et al. (2012) attempted to answer three research question by conducting experiments with the described architecture. They asked [*1*] how effective the novel prosody-based word extraction was, where *effectiveness* is measured as the percentage of meaningful words relative to the number of all words that were marked as salient. *Meaningful* words in this context are words that can be related to any dimension of the robot's sensorimotor data, in this case words that either label the object, or that relate to colour or size. Examples for *meaningful* words in this sense are thus "star", "red", or "large". Examples for *meaningless* words in this sense are "that", "good", but also "Deechee" because there is no *sm* dimension to which these words could be directly related to. [*2*] They further asked which *sm* dimensions would be associated with the salient words, and [*3*] whether word order could be derived from these associations, i.e. whether the received temporal order of salient words would be indicative of the standard word order in English - here *adjective - object word* rather than vice versa.

With regards to question [*1*] the analysis showed that a very high number of meaningful words were extracted, approximately 70-80% with each participant, and thus showed that participants did prosodically emphasise these words. Yet, notice that not all of the extracted words were meaningful.

With regards to questions [*2*] and [*3*] Saunders et. al. analysed both the difference in

information gain between the colour id and the object id dimensions of the *sm* data as well as the same difference between the size id and the object id dimensions. These differences were calculated for both tables. If participants primarily used the standard word order in English for 'modifications', e.g. "this is a red star", as opposed to the word order for 'predication', e.g. "this star is red", one would expect adjectives, if produced with at least average prosodic saliency, to be mainly located in the first tables. Object labels, on the other hand, under the same assumptions, would be primarily located in the second tables. This, in turn, would lead to an increase of the information index (or other correlation measures for that matter) of the size and colour dimensions of the *sm* data of the first tables and to a decrease of the information index of the object id dimension within the same tables. Conversely, one would expect to see in this case the opposite development in terms of changes of the information index in the second table if both 'modification' word order as well as the prosody criteria were met by the participants' productions.

The calculated differences then indeed indicated that these assumptions held in the case for the colour dimension but less so for the size dimension. The authors explain this difference with the circumstance that indeed not the ARToolkit shapes or tags attached to the boxes varied in size but that just the physical objects (boxes) did. This then had led to their participants not saying things like "this is a big red star", but to them producing descriptions such as "this is a red star on a big box".

Theoretical Approaches

A (so far) theoretical approach that fits in neither of the above categories but one which constitutes the only operationalisable framework for robots that, to the knowledge of the author, incorporates the notion of speech acts are the so called *semiotic schemas* of Roy (2005). Roy considers the grounding of language as a special form of the grounding of

an agent's belief, and further conceives of primitive speech acts as *intentional signs*. Yet he makes the somewhat irritating assertion that in their implemention "speech acts are assembled from lexical units (..) using a grammar"(Roy 2005, p.195). On the background that their semiotic schema framework supports precisely two types of speech acts, directives and descriptives, we take this to mean, that their grammatical parser recognizes the grammatical imperative in an utterance and subsequently flags the presence of a directive. If this is so, a single-word loud "No!" would go unnoticed. The learning and acquisition of these schemas is not part of (Roy 2005), which is a paper on representational issues.

2.5 Acquisition of Negation

Despite the majority of the first words of English-speaking toddlers being nouns, there are many other words in these toddlers' expressive vocabularies that do not refer to "anything visualisable" (Ryan 1974). *No* is one of these words. Due to the at least partially non-referential nature of these words their meaning or use can generally not be acquired exclusively by a process of association with an entity or event in the physical, non-social world. It is hard to imagine how they could possibly be taught via ostension, and often there is no indication that they would be used to name or label anything. The at least partially non-referential nature of these words in combination with the above mentioned focus of modern child language research on nouns and, to a lesser degree verbs, appears to effectively put these non-referential words and the process of their acquisition outside of the scope of most research on child language development.

This narrow focus may explain why there are comparatively few publications dedicated to the acquisition of negation most of which date back to the 1970's and 1980's. Another curious observation is the circumstance that many of the authors that did research on 'non-

referential' words and other 'non-referential' aspects of mother-infant conversations are in one way or another connected to Jerome Bruner, an early critic of the field's predominant focus on grammar of Chomsky's formalist programme and one of the most prominent figures in the first wave of the so-called sociopragmatic approaches to language acquisition (cf. Baldwin and Meyer 2007).

It is important to emphasise that the early emergence of "no" is not the only phenomenon that is outside of the scope of research that has its focus firmly on words that can potentially be explained by ostension or the co-occurrence of 'words and things'. Other early emerging words such as words to negotiate or answer questions ("yes") or words used within social practices such as greeting somebody ("hi"), saying good night ("night night") or saying good bye ("bye") which all belong to the very first words of an infant's expressive vocabulary cannot be explained by ostension or physical co-occurence of 'word and thing' (cf. Fenson et al. 1994, p. 93 for a list of the earliest words produced).

2.5.1 Taxonomies of early meanings of negation

A few taxonomies of early meanings of negation exist, yet we conceive of the differences between them as rather minor. Pea's (1980) taxonomy is the one that we chose as template for our taxonomies (cf. section 5.3.2). Pea gives an analysis of cognitive requirements of the various negation types, which is thought to explain the particular developmental order in which these various types emerge. These postulated cognitive requirements, affect, memory constraints, or the ability to perform logical judgements, then appeared to lend itself for computational implementation, which had been the initial reason to choose this taxonomy over the others. A second reason for the choice of Pea's taxonomy was the circumstance that it covers precisely the developmental period of our interest: the very beginning of toddlers' speech. The four children upon which Pea's observations and therefore his taxonomy are

based are between eight months and one year eight months old. The other two taxonomies of early forms of negation, which we considered, were the ones presented by Bloom (1970) and Choi (1988). Bloom's categorization is comparatively minimal as it only contains the three types *nonexistence*, *rejection*, and *denial*.

Choi's taxonomy is one of few negation taxonomies that we are aware of which is based on cross-linguistic data. Choi based her analysis on the observation of eleven children: two US-American, four Korean-speaking and five French-speaking children the latter of which contained one French-Canadian. She identifies nine different semantic/pragmatic types which are thought to emerge roughly in three subsequent developmental phases. It is important to note that the children under observation in Choi's study were slightly older than the ones observed by Pea. 'Choi's' children were between one year seven months and three years four months of age. According to Choi (1988) during the first phase the negation types *prohibition*, *rejection*, (*failure*), and (*nonexistence*) emerge, where the bracketed types have been observed to emerge one phase later with some children. In the second phase *denial*, (*inability*), and (*epistemic negation*) emerge, followed by *normative negation* and *inferential negation* in phase 3. We refer to (Förster et al. 2011) where we attempted to map Choi's types to the ones developed by Pea.

2.5.2 Pea's Taxonomy of Early Meanings of Negation

Pea's "taxonomy of negative meanings" (Pea 1980, p. 157) is depicted in figure 2.2. In his own words this taxonomy is "*not* a typology of early meanings for negation, for there is no reason to assume that these "types" .. are in any sense distinct in the children's conception of their own use of the negative words in these situations" (Pea 1980, p. 163). We take this, at least to us, slightly confusing emphasis on the taxonomy's elements being "meanings" but not "types of meanings" to mean the following:

Pea's taxonomy is very much motivated by the Wittgensteinian idea of family resemblance: "Like many words, "negation" does not have any one central or defining essence, but a number of meanings that partake of family resemblances to one another" (Pea 1980, p. 160: in reference to Wittgenstein 1958). This means that he does conceive of the various forms of negation not as 'concepts' that *all* share one or more essential features, but rather that *some* of these forms share certain common properties and might be used in similar situations for similar purposes. What appears to be equally Wittgensteinian' in spirit is the circumstance that the features which distinguish one type from another are pragmatic, or, in Pea's own words, "contexts for the use of negation".

At the topmost level the taxonomy has *adjacency* as criterion. Despite Pea's reason of this just being so "for the purpose of exposition", it emphasises the importance of the conversational context. Especially from a conversation analytical perspective the location of an utterance within a conversation is of utter import for its very function within that conversation. Yet the notion of adjacency that is used within the taxonomy is somewhat weaker than the notion of adjacency within the conversation analytical concept of adjacency pairs (cf. section 2.1.3).

On *Adjacency* In conversations the production of the first pair-part of an adjacency pair necessitates the production of a second-pair part by the addressee (cf. section 2.1.3). If for example Peter asks Susanne for the time, Susanne will feel urged to give Peter an adequate answer if possible. If she is unable to do so for Susanne might not have a watch, she will feel the need to explain why she is unable answer the question. Simple non-action on the part of Susanne would amount to a breach of conversational convention and would be considered odd or rude - in conversation analytical terms Susanne is accountable for her non-production. This is the strong version of adjacency, adjacency between parts of

adjacency pairs (cf. section 2.1.3). In Pea's taxonomy adjacency refers to the weaker, 'simple' conversational adjacency of a child's utterance to an adult utterance which naturally includes but is not restricted to second-pair parts of adjacency pairs.

For example the subscript of *truth-dependent denial* on the left hand side of the taxonomy-tree in figure 2.2 refers to "yes/no questions *and* declaratives". In the case of the child answering "no" in response to a truth-functional question such as "Is it raining outside?", the child is producing a second pair-part of the adjacency-pair *question-answer*. If, however, the "no" is adjacent to the declarative "It is raining outside", "no" is not a second pair-part, because *assertion-denial* is not an adjacency-pair. In the latter case the child could have just as well said nothing and would have not been accountable for its non-production.

Nonadjacency in this taxonomy simply means that it is the child that initiated the utterance. We take this to mean that also negative questions on the part of the child, if they would exist in the taxonomy, would be listed on the right-hand, non-adjacent side of the tree.

Study setup Before delving into the details of the different meanings of negation it is important to briefly explain some characteristics of the setup or situational context within which Pea's data was gathered. Pea's study is based on speech of toddlers gathered during monthly visits of 7 mother-child dyads during a single year. 5 of the children were aged 0:8 to 1:8 and the remaining two were aged 1:0 to 2:0, i.e. all children were in the so called one-word stage. During each visit the mother-child interaction during home activities such as playing, feeding, or bathing was observed for 90 minutes of which 30 minutes were videotaped for transcription purposes. Pea mentions additional audio recordings that were used for the transcription but he does not specify how much of the non-videotaped interaction

was (audio-) recorded, and based on which criteria the to-be-videotaped 30 minutes of the 90 minutes were selected . Importantly Pea, when counting 'acts of negation', also included gestural forms of negation such as head-shakes. These typically emerge ontogenetically before their linguistic counterparts. We would also like to emphasise the following shortcomings, possibly due to the absence of modern experimental standards 30 years ago, but which are important if we are to compare our data with the one presented by Pea. First, Pea does not specify the percentual proportion of audio- to video-recordings. Toddlers communicate frequently via gestures and via gesture-word combinations once they start to speak. Thus video-recordings should be preferred over audio-recordings. For this reason one would tend to be more suspicious about any categorization that is based on mere audio-recordings as gestures might have been used by the toddler that could substantially change the meaning of a word. Also, as we will see later on, facial expressions are of utmost importance if one seeks to determine the type of a negative utterance but also if one seeks to determine if such an utterance was meant seriously or not. The second major shortcoming is the lack of use of two or more coders such that we have to rely on Pea's judgement alone. Subsequently we don't know how reliable Pea's taxonomy is, nor do we know how reliable his frequency counts for the various 'types' are. Some of them could be based on very ambiguous decisions, and further one or more categories may have served as 'catch-all' categories similar to Dore's *practicing* category (cf. section 2.2.2, page 54).

The Frequent Five In this paragraph we will quickly summarize the five 'types', semantic categories, or meaning families which Pea observed to be the ones most frequently produced by his subjects. Also notice that some types occur on both sides of the 'adjacency-divide' such as *rejection* or *disappearance*. We list them roughly in their order of ontogenetic

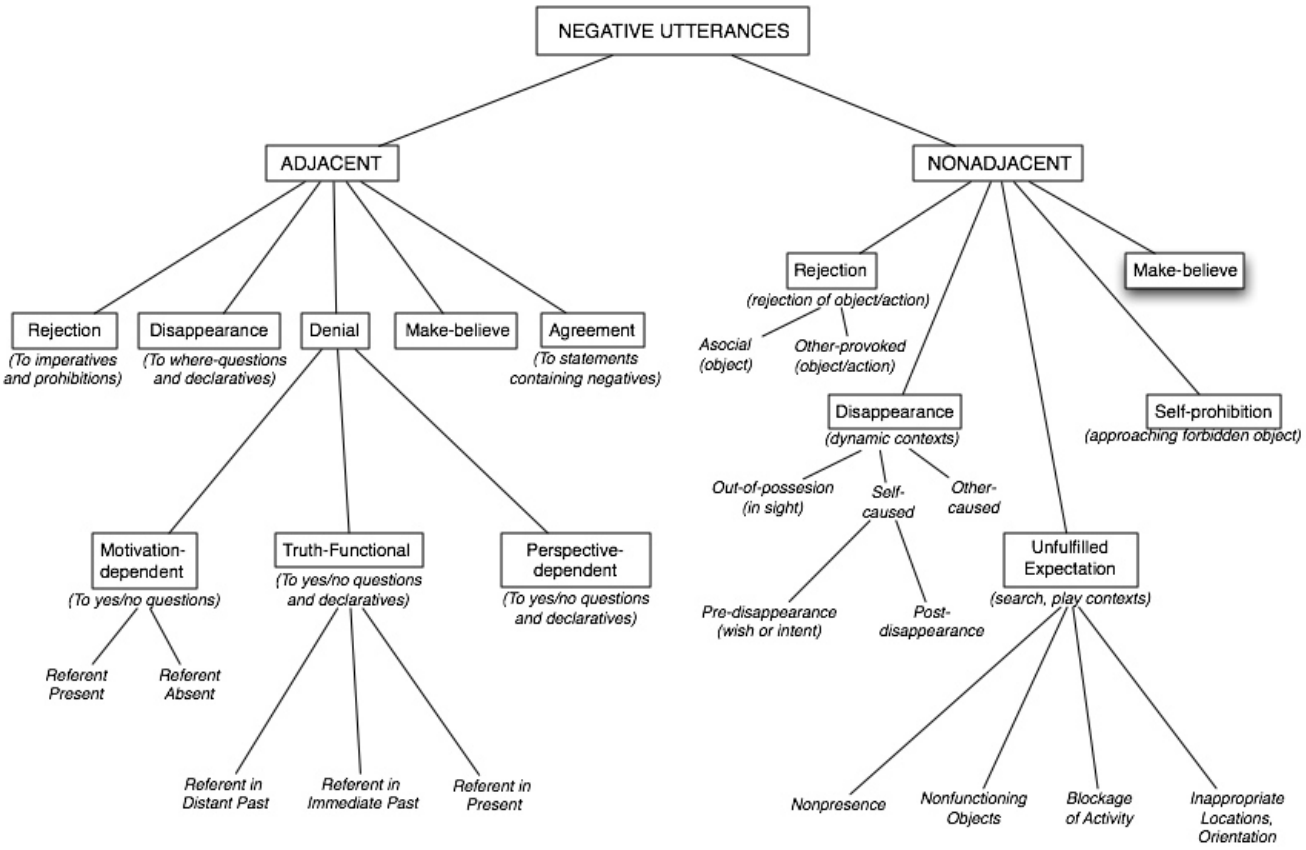


Figure 2.2: *Pea's taxonomy of "contexts for the use of negation", replicated from Pea (1980)*

emergence[11].

Rejection Pea defines *Rejection* negatives as those "action-based"[12] negations which are used by the child to reject events, persons, objects, or activities that are either in the immediate context of the child, i.e. the here and now, or which are "imminent in the mother's behaviour or utterance". Futhermore their "truth-value" is contingent on the child's motivation. Pea's definition of rejection seems to exhibit a certain overlap

[11] We say "roughly" because Pea reports some variation in the developmental order of *self-prohibition* and *unfulfilled expectation*. The remaining three 'types' emerged invariably in the given order.

[12] All citations in this and the following list items are taken from Pea (1980), including the conversational examples, unless marked otherwise.

with *motivation-dependent denials* as he also counts negative responses to desire questions such as "Do you want a cookie?" in this category. Pea invokes the response "no, now I do it" of one of his children to her mother's question "now can I do it?" as example for motivation-dependent denial. We don't see a fundamental difference between this response and a negative response to the foregoing cookie question[13]. Yet as rejection is also listed as nonadjacent "type" there are certainly rejections which are clearly not motivation-dependent denials. For example negative responses to non-linguistic behaviour such as mother's bodily indications of a looming nappy-change is a form of rejection that is not a (motivation-dependent) denial. The sole cognitive requirement for an agent to engage in this type is "the inner attitude of rejection" or "aversion".

This 'definition' of *rejection* appears to be largely identical to Bloom's version of *rejection*.

Self-prohibition *Self-prohibition* is "a form of egocentric symbol use in which the child approaches a previously forbidden object or begins to do something which has been prohibited in the past and then expresses a negative." It is somewhat of a misnomer in that *self-prohibition* does not mean that the child effectively ends up not doing what it is not supposed to do. The examples that Pea gives paint the picture of a child that shows signs of an internal struggle between ego and self, to use psychoanalytic terminology. So the child may approach a forbidden object, stop itself, possibly saying "no" while stopping, start a further approach, stop again, and so on. In the end the child might well end up having the forbidden object in its hand (or mouth) in case of the self winning the struggle. But what counts as self-prohibition are the expressed signs of prohibitive measures having

[13]For this reason we employed a clearer distinction between *rejection* and *motivation-dependent denial* in our own taxonomy which effectively renders our version of *rejection* a more narrowly defined 'type' than Pea's version of *rejection*

been internalized and being at work, the observation of a struggle that is not a struggle between mother and child but rather a struggle between the child and itself. In (Förster et al. 2011) we asserted that engagement in this type of negation had to be cognitively more complex as compared to rejection as the agent needed an internal representation to represent the preceeding paternal prohibition.

After having executed the experiment, we are far less certain, if this indeed must be the case. Our robot appeared to engage in self-prohibition within the prohibition scenario (cf. section 4.6) by producing words which were initially uttered in a prohibitive context by the respective participant. Sometimes, for example, the robot produced "can't", which clearly originated from the respective participant's use of "you can't touch that" or the like in previous sessions. We also observed some participants interpreting the robot's utterance as self-prohibitions, which was indicated by reactions such as "No, no, you can have that" with a prosodic emphasis on "can". As the robot's internal 'representations' are rather minimal and constant over the course of the experiment, the cognitive requirements needed in order to display something that appears to the observer to be some form of bodily and linguistic self-prohibition might not be so complex after all.

Disappearance *Disappearance negation* in Pea's taxonomy is a negative comment produced by the child such as "gone" or "no more" uttered upon the disappearance of objects or persons, the disappearance of something "which had been present just prior to the child's utterance" and denotes "the vanished object or event of disappearance". The engagement in this type of negation appears to require slighty more complex cognitive skills in the sense that the utterance refers to something that is not any more present in the immediate physical environment. Thus a child, according to Pea, needs at least some form of memory in order to notice the change between something just having been there

to something not being there any more. The disappearance may be self- or other-caused.

In (Förster et al. 2011) we re-asserted Pea's claim that a temporally somewhat extended memory would be needed in order to detect a *disappearance* event: one needs to 'know' that something was there in order to realize that it is gone. Based on observations made during the experiments we are now more critical about such stern claims. It is well possible that an agent might produce a negative word that *appears to be* a disappearance negative to an external observer without the aforementioned cognitive requirement being in place. Albeit the robot never produced disappearance negatives according to the coders, some of its words, which we termed 'pragmatic negatives' such as "go" (cf. section 5.3.3), at times appeared to have the potential to be interpreted as disappearance negatives. This is particularly so if such words are produced in the 'right kind' of situation, for example immediately after the child or robot had dropped something.

Unfulfilled Expectation *Unfulfilled Expectations* are somewhat similar to *disappearances* as they denote those uses of negation which refer to the "absence other than immediately prior disappearance or cessation", that is, the disappearance of objects on a longer time-scale than the immediate temporal context. Moreover negative utterances which refer to cessation events or, more generally, to "some aspects of the child's continuing line of activity (..) which does not occur" such as malfunctioning toys, or blocked movements of a bycicle, fall into this category of negative meaning. Through the inclusion of latter kinds of negative utterances this category differs from *disappearance* in ways other than a merely extended time-scale.

Bloom's category of *non-existence* seems to fall into this category of *unfulfilled expectations*.

Truth-functional denial *Truth-functional denial* is the last of the frequently produced negation types to emerge amongst 'Pea's' children and he deems this type to be "of a different logical order" ... " requiring abstract cognitive representations of yet greater complexity" than the earlier emerging types (Pea 1980, p. 166). As opposed to *rejections* or *motivation-dependent denials* this type refers to negative responses, typically a simple "no", to questions or declaratives which do not involve the child's motivation or affect in any form. For example a "no" that is uttered to answer the question "Did you crash my car?" is a form of truth-functional denial.

As the coders in our experiments deemed a great many utterances to be *truth-functional denials*, and granted that our robot does not engage in any form of logical judgement, this strong cognitive criteria might be slightly too strong. We will pick up this issue in our discussion in section 6.

Bloom's category of *denial* appears to be largely coextensive with Pea's *truth-functional denial*.

2.5.3 Affect as required 'skill' to engage in and acquire negation

Both Choi and Pea list *rejection* as earliest type of negation, although Choi also counts prohibition fully and nonexistence partially to the types emerging in phase 1. In Bloom's division, *nonexistence* precedes *rejection*, but Bloom seems to be the exception. This clearly indicates that, if we are to synthetically capacitate a robot to engage in negation, which it is ought to acquire in a developmentally sound fashion similar to the human acquisition process, the robot will need some form of affect or volition in order to "express inner attitudes of rejection" (Pea 1980, p. 165). This insight then forms the founding block for the symbol grounding system that is employed in this thesis.

However, granted that these affective requirements are given, the question as to how the linguistic skill to express negation comes about is still open. To this purpose two more ideas of how an agent, that 'has' some form of affect and is able to express it, could acquire this skill can both be found in (Pea 1980) and consitute the core of our two hypotheses (cf. section 1.1). Both ideas pertain to the interaction between a linguistically more skilled caretaker or teacher that holds more power within the interaction and a less skilled toddler or child.

Hypothesis 1 assumes that the conversationally dominant conversation partner will interpret emotional and volitional displays and gestures of the 'weaker' participant in a linguistic manner. This idea is based on the observation that caretakers of toddlers, typically mothers or fathers, have been observed to do precisely that: linguistically interpret their children's emotional and/or volitional states (Pea 1980, p. 179). In other words, the caretakers (or conversationally strong partners) produce words that fit the motivational or volitional state of their weaker partner. If this state is negative or rejective, the likelihood is very high that these words will be, at least partially, lexical negatives as in "No, no, don't like it".

We should emphasise that this potential "source of meaning" appears to be fundamentally different from ostensive sources of meaning: no joint attention, gaze following or mutual honing in on a referent is needed. What is needed, if an association mechanism is assumed, in order to establish the link between affect and word, is *simultaneity*: the word needs to be produced *while* the agent is in whatever state the interpretation 'refers to'. However, *simultaneity* is also required in ostensive theories of meaning, sociopragmatic or else, if simple association as learing mechanism is assumed: the relevant words have to be uttered while joint attention is established, or while the joint action is executed. They must not be produced before or after the time windows during which this is the case.

It is surprising that the phenomenon of intent interpretation is hardly ever mentioned in standard textbooks on language acquisition and word learning, and that we had to 'dig deep' to find it documented in the literature. This is even more surprising once one has observed this social mechanism in vivo: it appears to be the most natural kind of behaviour between a caretaker and a child.

Hypothesis 2 assumes that caretakers employ negative words, when physically prohibiting a child (Pea 1980, p. 181: referring to Spitz 1957). Again simultaneity is required if the, often implicit, assumption is made that the learning process works via association: The negative word of the caretaker has to co-occur with the negative motivational state, that is brought about by the caretakers intervention. We will see in our analysis that this assumption might be problematic.

I envy you the certitude of your grasp of the causal well-springs of human behavior. It is apparently quite clear to you that you drive on the left in England *because* there is a rule which tells you to; you apparently have been able to reject quite firmly some not unrelated possibilities, such as that you are oriented to the possibility that other drivers will be oriented to the rule, and that if they (and you) do otherwise you are likely to collide head on, it being the avoidance of this prospect which motivates your compliance, rather than "because it is a rule".

—*Emanuel A. Schegloff (in response to John R. Searle)*

Chapter 3

A Robotic Architecture for the Acquisition of Negation

Conceptually the architecture that is presented in this chapter, and which was employed in the experiments described in chapter 4, is based on the system outlined in (Saunders, Nehaniv and Lyon 2011)[1] which has been used successfully to acquire object labels and adjectives from dialogues with naïve participants. In terms of the actual software all modules but one were re-developed in order to integrate new capacities of the iCub that had been developed within the RobotCub and ITALK projects since then (RobotCub project 2013, ITALK project 2013).

Figure 3.1 gives a functional overview of the system architecture. Note that the functional modules, depicted with boxes, do not necessarily coincide with software modules in the repository. The only truly new module, as compared to Saunders' system, is the *motivation system*. Furthermore the *body behaviour system* differs greatly from the equivalent system employed by Saunders. This circumstance is owed to the mentioned new

[1] We will refer to this system in the following as *Saunders' system* for ease of reference. No denial of the other authors' merits is intended.

capabilities of the iCub platform such as force control. Apart from these, the need for new behaviours such as *reaching* or *rejecting*, and the way behaviours are triggered necessitated a complete reimplementation of the way the humanoid behaves in any particular situation. All modules are subsequently described in more detail.

3.1 Perception System

The perception system processes camera images from one or two cameras and provides the other modules with high-level information about salient objects or actions. Furthermore low-level percepts from the motor encoders, external pressure excerted upon the arms, and the detection of *pickup* and *put-down* events of salient objects can be propagated into the system if needed (see below, this section). The detection of external resistance is only available when the arms are operated in force control mode and is currently only used in the prohibitive scenario described in section 4.6.

The core element of the perception system is a modified version of a salience module originally developed by (Rüsch et al. 2008). Amongst other changes additional salience filters for a commercial face tracking software (faceAPI 2013) as well as for ARToolKit tags (ARToolKit 2012) were integrated into the module. The ARToolKit software package provides for comparatively easy object recognition by means of the attachment of easy to recognize, adaptable, black-and-white symbol tags to target objects. It was used in order to avoid common computer-vision problems that typically emerge in the context of a reliable object detection. In the future this system should be replaced by a "tag-less" system. Despite the availability of many salience filters, such as skin colour or movement detection, that were inherited from the original salience module, only the two abovementioned ones were used for our experiments (cf. chapter 4).

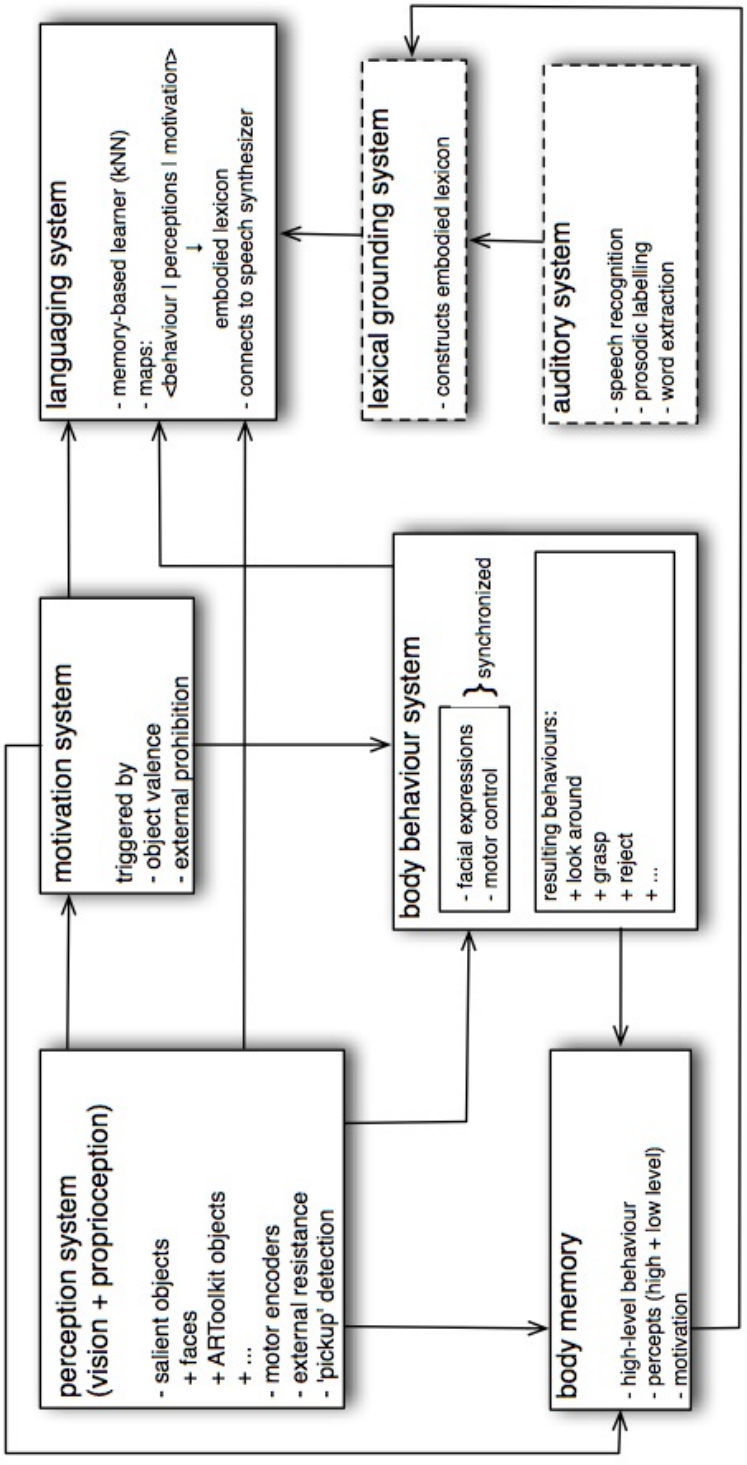


Figure 3.1: ***Functional overview of robotic architecture for language acquisition****. Solid lines indicate components that are active during experimental sessions ("online"), dotted lines indicate components that work offline.*

As, within these experiments, typically more than one object is located within the visual field of the robot, an algorithm was implemented that detects when an object is picked up by the participant and when the object is put back onto the table. We will call these events *pickup* and *put-down* in the following. This algorithm makes use of an estimation of the z coordinate ('vertical axis') of all objects in the Cartesian space of the iCub's root reference frame (cf. iCub Forward Kinematics 2012). *Pickup* and *put-down* events are crucial as they trigger different behaviours (cf. section 3.3), which in turn trigger speech actions in the languaging system (cf. section 3.4). Thus, via picking up objects and presenting them to the robot, participants drive the interaction via said trigger events without necessarily being aware of it.

3.2 Motivation System

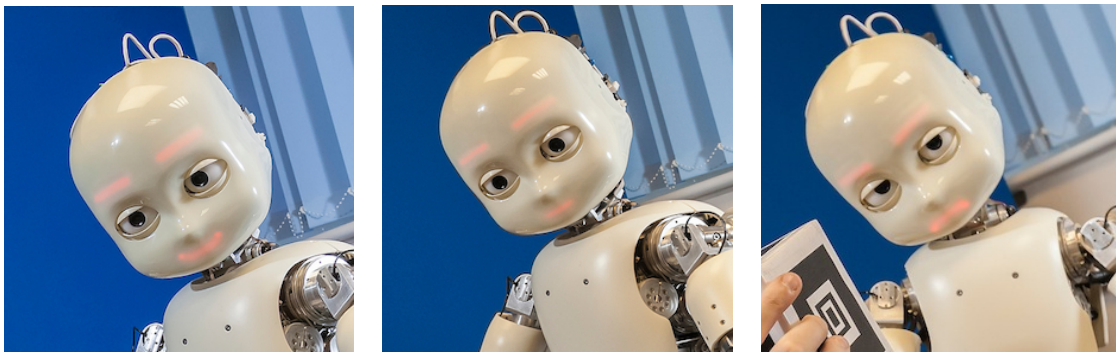
The motivation system provides the system with an additional dimension for the sensorimotor data vectors and is treated in exactly the same way as the other dimensions of said vector. The motivation model is deliberately kept as minimal as possible in order to investigate the dynamics between language acquisition process and motivation. The motivational state is a numerical value between -1 and 1 and is currently discretised within the lexical grounding system. This means that from within the latter system only three motivational states are distinguished: -1 (negative), 0 (neutral), and 1 (positive).

In the experiments described in the following chapter, the motivational state of the robot is neutral unless the participant picks up an objects that has a negative or positive valence. The *object : valence* mapping can be either generated randomly or can be specified in a configuration file. In the prohibition scenario the motivation is furthermore modulated by external resistance on the right arm. Through the manner in which said scenario is set

up, detecting external resistance implies that the participant is physically restraining the robots arm movement. If such an event is registered the motivation value is set to negative for a certain specifiable time frame.

3.3 Body Behaviour System

The body behaviour system generates the humanoid's bodily behaviour including its facial expressions. It receives input from both the perception and the motivation system. The central design tenet was to make the robot act as believable as possible, as opposed to making its movements as accurate as possible. A consequence of this tenet is that the system should always produce some kind of behaviour. Toddlers don't freeze and neither should the robot. This means for example that the robot continues to behave even if the majority of its perceptual capabilities fail and the available object coordinates are out-of-date. Naturally there is a limit to this design target: if *all* perceptual capabilities fail long enough, the robot will stop behaving as all of its behaviour is target-directed with the target being either an object or a human face.



(a) Happy *facial expression* **(b)** Neutral *facial expression* **(c)** Grumpy *facial expression*

Figure 3.2: ***Facial expression as used within the experiments.*** *(Images are clippings of photos taken by Pete Stevens)*

The author put a considerable effort into the fine-tuning of the system until it roughly generated the kind of behaviour which he felt would be believable enough for it to be used within the experiments. We feel incapable of describing this process in an adequate textual form, a process not dissimilar to the "black art" of designing a neural network which involves, amongst other things, choosing a particular input and output coding, choosing the number of network layers, and choosing a particular learning algorithm, which, in turn, typically necessitates the choice of some learning parameters. The kind of fine-tuning necessary for our behavioural system was mainly related to the adjustment of time constants that impact on the duration and synchronization of the described behaviours as well as the coordination of sub-actions that are constitutive for some of the behaviours.

An important example for time constants involved in the temporal coordination of sub-actions is the duration of the robot's gaze at the human face in relation to the duration of its gaze at objects, both of which are constitutive components of the *Watching* behaviour. In human face-to-face interaction the difference between a stare and a casual glance is mainly constituted by the duration of the gaze. It is therefore hard to over-emphasise the importance of the various time constants that ensued from this process. We refer the technically oriented reader to the software itself, its documentation, and the configuration file where these constants are specified. (cf. `<DVD>/software/italk/src/negationBehaviour/` and `<DVD>/software/italk/app/negation/conf/negationBehaviour.ini`).

Five different behaviours have been implemented:

- Idle
- Looking around
- Reaching for object
- Rejecting
- Watching

Each behaviour has an associated unique behaviour id which is broadcasted to the other subsystems whenever a change of behaviour occurs. A sketch of the 'behavioural loop' of the body behavior, which outlines when which behaviour is executed, is given in algorithm 1, followed by a description of each behaviour.

3.3.1 Behaviours

Idle

As the name indicates the robot does not behave in a meaningful way while being in this state. This behaviour is currently only activated during an initial calibration phase upon startup of the system while connections to controllers and other modules are established. The facial expression during this behaviour is neutral.

Algorithm 1 Outline of the 'behavioural loop' for the robot's body behaviour. Notice that 'offer_detected()' is based on information broadcasted by the perception system, and 'valence()' is based on information pertaining to the motivation system.

```
1: while negation behaviour module is running do
2:   if ! headController→connected() then
3:     behaviour = IDLE
4:   else
5:     if ! offer_detected() then
6:       behaviour = LOOK_AROUND
7:     else
8:       getObjectID(oid)
9:       if valence(oid) > neutralThreshold then
10:        behaviour = REACH_FOR_OBJECT(oid)
11:      else if valence(oid) < -neutralThreshold then
12:        REJECT(oid)
13:      else
14:        WATCH(oid)
15:      end if
16:    end if
17:  end if
18: end while
```

Looking around

This behaviour is the default behaviour after the initial calibration phase. The iCub switches its focus between the different available objects in its visual field and the human's face, if the face tracker can detect one. If the face tracker fails to detect a face, a default face location is assumed based on the spatial position of the sitting participant relative to the robot. The facial expression is neutral. The amount of time that if focuses on each object as well as the time it focuses on the human's face are adjustable via configuration file. The particular trajectories of the head-eye movements are calculated and executed by a kinematic gaze controller, developed by Pattacini (2010). This controller was designed to generate biologically plausible, human-like combined head and eye movements.

This behaviour can be observed in any of the experiment-related video recordings contained on the appended DVD. It is easiest to spot at the very start of the interactions before participants pick up the first object.

Reaching for object

This behaviour is executed when a participant picks up an object with positive valence, thereby triggering positive motivation. The facial expression is *happy*. After a short (adjustable) time of looking at the object, the iCub reaches out for the object. The palm of the hand is facing upwards in order to signal to the participant that he or she can put the object into its hand.

We initially considered having the robot grasp directly for the objects. Due to the lack of a reliable grasping module at the time of implementation but also due to the size of the objects, we decided against this behaviour. Moreover, upon starting the first experiment, it quickly became clear that the open-hand gesture is immediately picked up by participants and possibly even favourable over the more 'aggressive' direct grasp. We conjecture that it

might be preferable to the direct grasp as it has a gestural (and therefore communicative) quality that the direct grasp may have not. It would be of no surprise to us if the open-hand gesture might have contributed to a higher degree of involvement of participants with the robot. Our impression is, that, possibly caused through conventions of politeness, and the imperative nature of the gesture, it appears to urge participants to physically interact with the robot. It further may have contributed to an increased ascription of intentionality to the robot on the part of the participants. We did not formally test the suspected contribution of this behaviour to participants' involvement nor did we test for the degree of ascribed intentionality[2]. Nevertheless we wanted to bring its potential importance to the reader's attention.

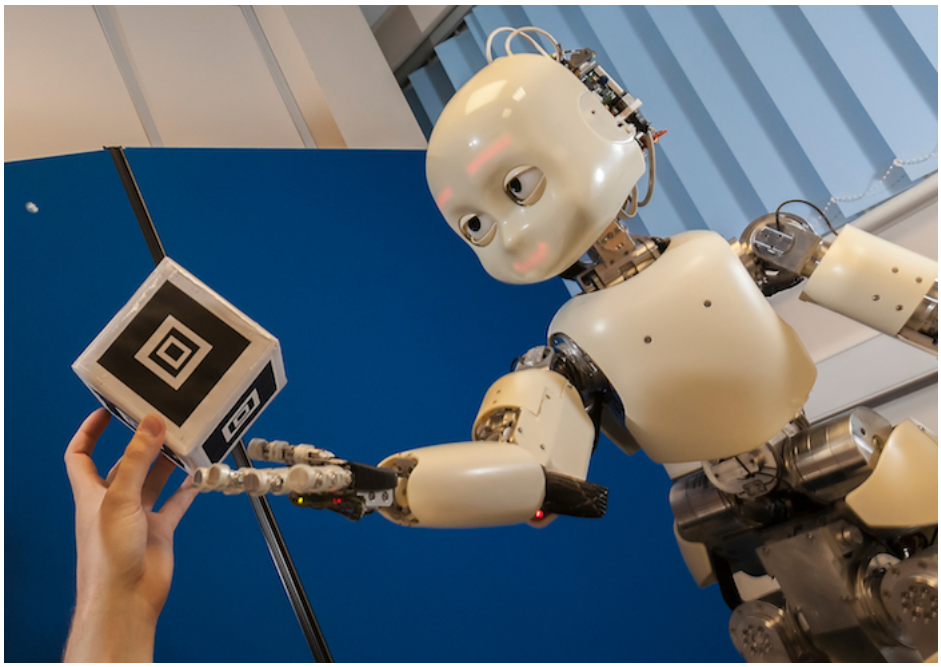


Figure 3.3: ***Reaching with open-hand gesture****. (Photo by Pete Stevens)*

[2]We only became aware of the existence of psychological tests for ascribed intentionality long after the experiments had ended.

While reaching out, the iCub's focus of gaze switches between the selected object and the participant's face. The duration of both gaze actions is adjustable and was fine-tuned. The reaching movement is executed via the cartesian controllers of the arms (cf. Pattacini et al. 2010). Due to a certain amount of noise in the distance estimation of the object, potential occlusion of the frontal side of the objects when in the robot's hand, and unpredictable detection-switches between various sides of an objects[3], this in theory uniform behaviour manifests itself as a variety of similar but non-repetitive observable behaviours (cf. figure 3.3 for one example). These behaviours include the dropping of an object, the movement of an object closer to the body, the giving-back of an object to the participant, the holding of an object above the head. None of the latter behaviours are part of the control design, they 'emerge' mostly due to the mentioned forms of noise, and, to a lesser degree, due to the particular way a participant interacts with the humanoid, i.e. the way he or she places an object into its hand, the way he or she holds the object, potentially causing the occlusion of tag markers, etc. Observations during the experiments, comments made by participants, and even comments of participants that were directed towards the humanoid, indicate that this particular behaviour often invokes the impression of intentionality on part of the robot - intentionality by noise, so to speak.

Prohibition Scenario Within the prohibition scenario, which will be introduced in section 4, participants were asked to and taught how to restrain the robot's arm movement as soon as it tries to approach a forbidden object. Within this scenario the arm controllers are operated in a novel way termed *force control*, a different control mode as compared to the one in operation within the rejection scenario. This control mode makes it possible

[3]All presented objects were cubes with identical ARToolKit tags attached to every side of the cube. For this reason often more than one tag was visible to the robot's camera and the object detection was observed to switch its 'focus' at times between the visible tags.

that the robot can actively move its arms while, simultaneously, external pressure might be applied such that the arm moves in a compliant manner. If external pressure is detected, which in our scenario equates to physical restraint of the robot's arm movement, the robot's motivational state is switched immediately to negative and a *grumpy* face is displayed. The design rationale behind this 'motivation switch' is the idea that agents don't like their agency to be diminished (cf. hypothesis 2 in section 1.1). The reaching behaviour is in this case *not* aborted. The decision not to abort the reaching, despite participants' attempts to counteract the robot's movement (which they were told to do), was made in order to simulate a behaviour akin to a toddler's insistence to engage in a forbidden act. Furthermore, the reaching behaviour was aborted frequently due to the object recognition losing track of the object. Unplanned abortions of this behaviour were indeed the norm rather than the exception without any further additions or checks added to the control loop - they are, in some way, emergent features of the interaction caused by the kind of aforementioned noise. It is only in the prohibition scenario and only in the just described situation, that the reaching behaviour may co-occur with a negative motivational state (cf. figure 3.4).

The reaching behaviour can be observed in any experiment-related videos of any session on the appended DVD. The prohibition task, i.e. participants physically restraining the robot's arm movement, was in place during the first three sessions of the prohibition experiment (participants P13 - P22). Not all participants followed our instructions and restrained the robot's arm in the respective situations. Good examples of the ensuing interaction of those who did, can be seen with participants P15 and P16.

Rejecting

The *Rejecting* behaviour is executed when a participant picks up an object with negative

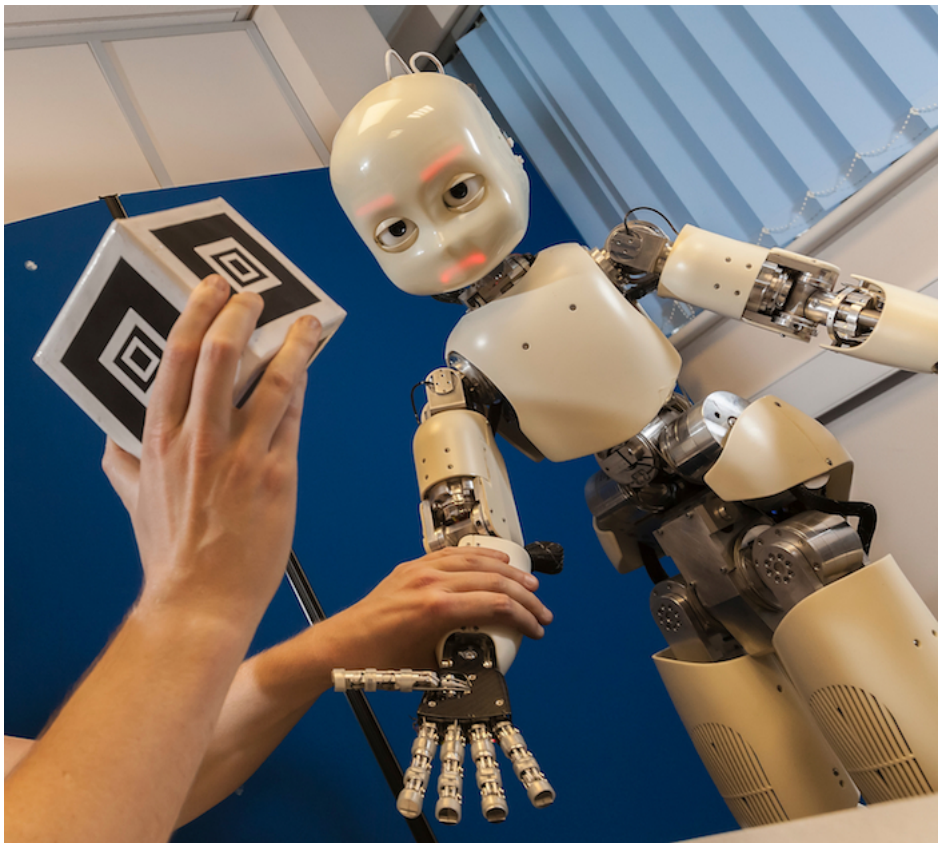


Figure 3.4: ***Physical restraint of the arm****: participants were taught, how to restrain the iCub's arm movement if it approaches a forbidden object. (Photo by Pete Stevens)*

valence, which triggers negative motivation. After a short (adjustable) time of looking at the object the iCub starts to frown, facial expression *grumpy*, and turns its head away. The particular way in which the head is turned depends on the particular position of the object relative to the right camera/eye position. Depending on the centre coordinates of the object, in 2D camera image coordinates, an avoidance vector is calculated that roughly points in the opposite direction of the vector $<image\_centre> \rightarrow <object\_centre>$. We say *roughly* because the horizontal component of the vector is emphasized by multiplying it

with a scaling factor in order to pronounce the yaw movement (looking sideways) over the pitch movement (looking down). Figure 3.5 depicts the relevant vectors, and algorithm 2 specifies the calculation in more detail.

Within the experiments *xSkew* was set to 3, which was determined experimentally by trial and error. Due to slight movements of the object in the participant's hand in combination with previous head movements of the robot, the results are typically dynamic,

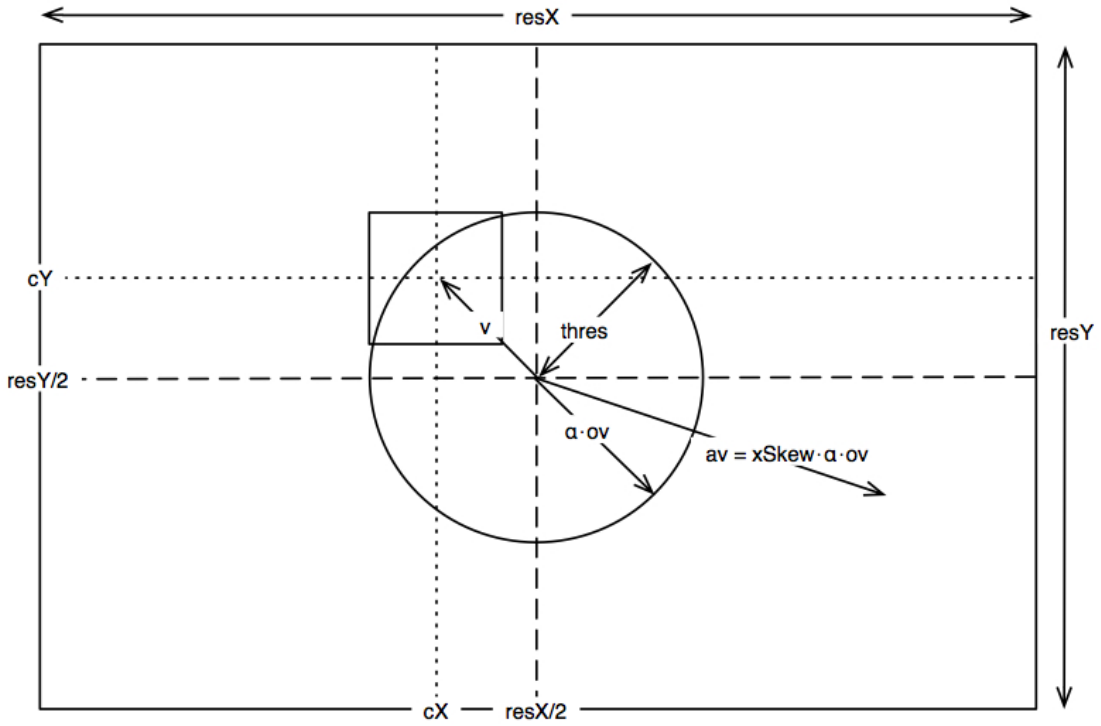


Figure 3.5: ***Calculation of avoidance vector.*** *The object, represented by the square as determined by the perceptual system, is located in the upper right, but with its centre still within the inner region of the robot's visual field. The avoidance vector (av) is calculated as indicated, pointing roughly in the opposite direction of the object and is subsequently sent to the head controller. v: centre of object,* resX*: image width in pixels,* resY*: image height in pixels,* thres*: radius of centre region within which the avoidance is active,* $\alpha \cdot ov$*: scaled "opposite vector", av: avoidance vector,* xSkew*: "skew" factor to emphasise the x-component of av. Notice: this equation is mathematically incorrect and is meant to indicate that xSkew only operates on the x-component of the ov-vector. See algorithm 2 for the correct formulation.*

Algorithm 2 Object-gaze-avoidance: Determination of target coordinate for head-eye movement in order to avoid looking at object; t_a is adjustable via configuration file

1: Fixate on presented object $o = \binom{cX}{cY}$ for t_a seconds
2: **if** valence(o) < 0 **then**
3: **if** $|\binom{cX}{cY} - \binom{resX/2}{resY/2}| < thres$ **then**
4: $ov = -v = \binom{resX/2-cX}{resY/2-cY}$ {Calc. "opposite vector" ov}
5: $av = \alpha \cdot \binom{xSkew \cdot ov[0]}{ov[1]}$ {Calc. avoidance vector}
6: Send $\binom{resX/2}{resY/2} + av$ to head controller
7: **end if**
8: **end if**

believable, and somewhat unpredictable head movements such that the iCub looks away from the presented object. Sometimes this results in what some participants referred to as head shaking, which one of the participants, P04, described at some point as creepy. The arms are not moved during this behaviour.

The rejection behaviour can be observed in all videos of the experiment. For visual examples for the occasionally emerging 'head shakes', and participants' interpretation of them as such see section D.4.1.

Watching

The *Watching* behaviour is executed when a human participant picks up an object with neutral valence, which triggers neutral motivation. The iCub switches focus between the selected object and the participant's face. The duration of both object- and face-directed gaze is adjustable, typically not identical, and was determined experimentally. The facial expression is *neutral*. The arms are not moved.

The watching behaviour can be observed in any of the experimental videos on the appended DVD.

3.4 Languaging System

The general purpose of the *languaging system* is to produce utterances, based on what participants said in previous sessions, and which are meaningful in the particular situational context. To this purpose it uses the *embodied lexicon* which is generated offline by the *lexical grounding system* (cf. section 3.7) and takes into account the sensorimotor-motivational data provided by the following other subsystems:

- The *body behaviour system* provides information about the currently executed behaviours in the form of behaviour ids.
- The *perception system* provides information about the current perceptions in the format of a vector of sensorimotor data from the motor encoders, higher-level vision percepts on recognized objects, a binary value from the face-tracker (face recognized or not), and potentially a binary value indicating the experience of external resistance.
- The *motivation system* provides information about the current motivational state in the form of a single motivation value.

The information of these three subsystems is consequently combined into one vector which will henceforth be called the sensorimotor-motivational (*smm*) vector (cf. figure 3.6). In short, this vector contains all the information about the humanoid's own bodily and motivational state, the behaviour that it currently executes and all perceptions related to its environment that it has access to (currently only ARToolKit objects, and human faces).

Note that for the experiments described in chapter 4 the decision was made to remove those dimensions from the smm vector that originate from the motor encoders in order not to fall victim to the curse of dimensionality (cf. Hastie et al. 2013, section 2.5).

One may criticise that we thereby render the learning problem unduly easy by throwing

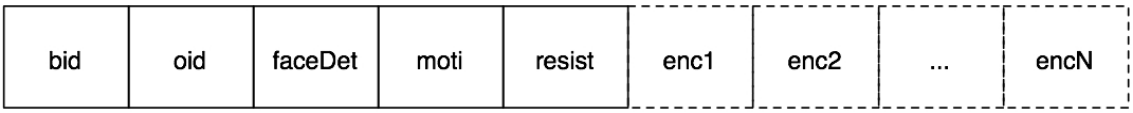


Figure 3.6: ***Sensorimotor-motivational (*smm*) vector****. Solid line indicates dimensions that were used within experiments;* bid*: behaviour id,* oid*: object id,* faceDet*: face detected,* moti*: motivation value,* resist*: resistance detected,* encX*: encoder #X,*

out a majority of bothersome data. By only keeping potentially meaningful dimensions, we indeed incorporate external (designer) knowledge, which a child may not have. Yet, consider the *whole-object constraint* mentioned in section 2.2, which may be cited in defense of this decision. Based on this and other constraints one could argue that the child may have said 'designer' knowledge, brought about by some evolutionary process and encoded in its genes. Nevertheless we do agree that, in principle, one should keep such incorporation of knowledge as minimal as possible. Yet the critic has to consider that our and Saunders, Nehaniv and Lyon's (2011) approach are the first of its kind in that it combines lexical entities extracted from a linguistically unconstrained human-robot conversation with a learning system. We could not possibly have performed any algorithmic tests, to choose and adjust algorithms to the given learning problem due to the lack of a realistic data set. We neither knew from the outset how the word distribution in the embodied lexicon would look like, nor did we have an estimate of the ratio between 'true positives', i.e. negative words associated with *smm* data with a negative entry in the motivation dimension, and 'false positives', i.e negative words with an 'attached' positive motivation. In other words, we were faced with an unknown level of what might be conceived of as noise in our data set. If the noise surpassed a certain level, this would deteriorate the performance of any learning algorithm, associative or other. The knowledge required to generate such a synthetic data set, i.e. how often, with what prosodic characteristics, and at what time within the

conversation, participants would utter which words, is only, tendentially, available now after we actually executed the experiment. Even more important, the human effort necessary to conduct these experiments was by several magnitudes larger than the one required to solve standard machine learning problems such as image recognition, classification of medical data, email spam etc., and was caused by the semi-automatic speech processing (cf. chapter 4). Thus, our reply to criticism, that we did cut corners, by pre-selecting certain dimensions from our robot's embodiment, is, that we simply could not take the risk of choosing a potentially non-efficacious algorithmic setup due to the fact that we could run this experiment with one such setup exactly once.

The central element of the *languaging system* is a memory-based learner. In particular, the system makes use of an efficient implementation of the k-nearest-neighbour (*knn*) algorithm, the open source Tilburg Memory-Based Learner (TiMBL 2012, also cf. Daelemans and van den Bosch 2005). This choice was made, simply because (Saunders, Nehaniv and Lyon 2011) used this algorithm successfully before in the context of noun acquisition, and because memory-based learners do not require a training phase as many other machine learning algorithms. In other words, they are a very easy to use family of algorithms with a very 'low maintenance' due to the non-explicit learning phase. The author has no strong conviction that this type of learning algorithm is superior to other machine learning techniques on this particular kind of data.

For the experiments described in chapter 4 the parameter k, the number of neighbours, is set to 3. As opposed to (Saunders, Nehaniv and Lyon 2011) we used as distance metric the modified value difference metric (*MVDM*) after initial experimentation with different parameters and upon recommendation by one of the authors of TiMBL (personal communication). For a description of the k-nearest-neighbour algorithm in combination with the MVDM metric see Cost and Salzberg (1993). For the experiments described in chapter 4

we employ something that we call a *differential lexicon* (see following subsection).

This lexicon is simply a multitude of embodied lexica, each of which has one word removed as compared to the base lexicon from which the set is constructed. The reason for having a multitude of lexica instead of having just one, is the lacking support of TiMBL to temporarily suppress single items once the data base, in our case the embodied lexicon, is loaded.

During test runs we observed that the use of a single embodied lexicon often resulted in the repetitive sequential production of the same word, presumably due to a lack of change within the *smm* vector. We therefore decided to prevent repetition by temporarily suppressing the previously produced word. Thereby the variety of the linguistic behaviour increases without the need on part of the designer to make any biased decision as to what kind of lexical items would be eligible for production after a particular (type of) item was uttered.

The kind of biased decision that we have in mind, is external linguistic knowledge involving things such as grammatical word order, grammatical scope of a negative operator, or lexical/semantic type. A tempting incorporation of such knowledge for example, falling into the category of aforementioned word-order constraints, would be the temporary deactivation of adjectival[4] items once they were uttered, in order to increase the likelihood of a noun being produced next. An increase of adjective-noun constructions could the outcome of such incorporation of designer knowledge.

[4]How we would determine, what is adjectival, is a separate question. In our system this could be achieved by adding a pre-processing step, in which a hand-constructed set of 'semantic adjectival prototypes' could be test-fired against the embodied lexicon, in order to extract and mark potential adjectives. Of course the same could be done for nouns, emotion words, verbs, etc.

3.4.1 Differential Lexicon

The differential lexicon consists of the k-nearest-neighbour implementation and a set of embodied lexica, i.e. lexica containing grounded words. It is created each time anew upon startup of the system based on the (single) embodied lexicon created by the *lexical grounding system* (cf. section 3.7). The lexicon determines which lexical item best matches the current embodied and situational context, by comparing the current *smm* vector of the robot with the *smm*-parts of the grounded words of the active lexicon. On startup a separate memory base m_{lex_i} for each distinct lexical item lex_i of the *embodied lexicon* is created that contains all entries of said lexicon except of those whose word-part matches the one of lex_i. Furthermore one full memory base m is constructed that contains all the words from the *embodied lexicon* and which is therefore identical to this lexicon. Hence the system constructs $n + 1$ memory bases for an *embodied lexicon* with n distinct words.

Each m_{lex_i} as well as m serves as a data base for the *knn* algorithm. As TiMBL does not allow switching the underlying data base during run-time, $n+1$ instances of the *knn*-learner are created, one for each memory base. We will henceforth abbreviate that instance of knn, which operates on the full embodied lexicon, knn_{full}, and those instances that operate on derived lexica, missing a particular lexical item lex_i, knn_{lex_i}. For each instance of *knn* the *gain ratio* (cf. Quinlan 1992) between the lexical items and each other dimension of the particular memory base is calculated. This effectively constitutes a measure of *mutual information* between each *smm* dimension and the dimension that holds the lexical items[5]. The resulting gain ratios are subsequently used by the *knn* algorithm as weights, associated with each *smm* dimension, that are used in order to calculate the distance between a new, yet-to-be-classified data point and the data points in the particular memory base.

[5]Cf. Daelemans and van den Bosch (2005) for a more detailed discussion about different measures to calculate the relevance of dimensions in memory-based reasoning.

The *best match* or *winner* is determined in terms of the smallest distance between the current, to-be-matched *smm* vector and the *smm* parts of all entries in the active memory base (see next section on the activation of memory bases).

3.4.2 Operational Description

An important question, that has not been touched so far, is as to when the humanoid actually speaks. Within a dialogue, humans follow various cues such as prosodic marking of questions, pauses, and non-linguistic cues such as particular properties of the interlocutor's gaze in order to regulate turn-taking behaviour (cf. section 2.1.3). More generally, if spoken language is regarded as a form of social action it becomes one of several goal-driven actions of an agent. Taking this perspective leads to the question of when an agent would speak in the first place as opposed to choosing a non-linguistic action. A follow-up problem, in case that the decision was made in favour of speaking, is the issue when precisely the robot is supposed to speak, when does it "think" that it is its turn.

In our experimental scenarios the first problem, i.e. when does the robot speak at all, is 'solved by design' as the particular situation in which the humanoid performs linguistic actions is invariably a teaching scenario: the humanoid *reacts* to the participant's action rather than proactively initiating a dialogue. Furthermore, it always reacts this way, that is, it never runs away[6].

The solution on how to operationalise turn taking, i.e. the solution to the second problem, is solved rather trivially, and, more importantly, unsocially. This means that a proper solution to this problem has yet to be found. We will see the outcome of this lack of a proper solution later in the analysis of the experiments (cf. chapter 5) and we will discuss this issue in chapter 6. As, to the knowledge of the author, no operational models

[6]If a speech act is considered to be one of a variety of possible (re-)actions in a given situation, the action of 'running away' could be amongst the set of possible reactions when being asked a question.

of turn-taking for robots exist, the decision was made to use a mechanism akin to the one employed by Saunders, Nehaniv and Lyon (2011). This means that the differential lexicon is activated whenever the system receives the notification from the *body behaviour system* that a trigger behaviour is being executed. The experimenter can define which behaviours are trigger behaviours via a configuration file. In our experiments *Watching*, *Rejecting*, and *Reaching for object* are trigger behaviours.

Upon activation knn_{full} is executed and returns the best matching lexical item. knn_{full} is executed as often as the system receives inputs from the perception system. The frequency of the perception system sending out new percepts is limited through the time needed to compute high-level percepts from the vision system, some of which are computed in parallel. This results in a frequency of approximately 30 Hertz[7]. In other words, the robot queries its lexicon constantly as to which of its entries fits best to the robot's current situation in terms of high-level percepts, motivation, etc.

As it is neither sensible nor feasible to have the robot speak with approximately 30 Hertz, i.e. every 33 milliseconds, and in order to stabilize the system with regard to potentially erroneous detections of the vision system, the following thresholding mechanism similar to the one used by Saunders, Nehaniv and Lyon (2011) was introduced.

Three different thresholds were chosen for the three motivational states of the robot: positive motivation has assigned the lowest threshold (highest probability of speaking), neutral motivation has assigned the highest threshold (more reluctant to speak), and the assigned threshold for negative motivation was chosen to be in between the two former thresholds.

The design rationale behind these choices follows the intuition that we speak about things that we like (positive motivation) rather than things we don't care about (neutral

[7]The perception system could certainly be optimized in terms of the computational complexity of the vision part, but test experiments showed that said frequency is sufficient for the experimental scenarios.

motivation), and that we speak in order to stop actions that we dislike (negative motivation) rather than if we don't care.

For each lexical item a counter is held that is increased by some increment in case of this item being a best match. Said counter is decreased by some different increment if this item is not the best match with 0 being the lower bound. A further increment is added to a winning item's counter in case this item has also been the winner of the previous match. All three different increments were chosen experimentally during test-runs of the system such that the robot's production frequency roughly matched our intuition.

As soon as a lexical item lex_i reaches the motivation-dependent threshold the word-part of the lex_i is sent to the speech synthesizer and the iCub subsequently utters this word. Upon speaking, all counters are reset to 0, knn_{full} is deactivated, knn_{lex_i} becomes active, and the matching cycle is repeated. As soon as the value of any dimension in the *smm* vector changes knn_{full} is activated again. The only dimension that is not considered in terms of detecting changes of the *smm* vector is the dimension that represents the signal from the face tracker. The latter proved to be volatile to the degree that its inclusion would have jeopardized the intended operation of the differential lexicon as a whole.

3.4.3 Non-technical summary of the languaging system

In non-technical terminology, the questions as to when the robot speaks, how often the robot speaks, and what it says, may be summarized as follows: The robot speaks whenever a participant picks up a box and presents it to the robot, granted that this participant has spoken to it before, which means that the robot remembers some of the words, that the participant used previously as well as the situations in which he or she used them. As soon as the robot is certain enough, that a word that it remembers, matches the particular situation in terms of its own motivation, in terms of its own behaviour, in terms of the

presented object, and, in case of the prohibition scenario, in terms of whether its arm is restrained by the participant or not, it says the word. The reason for the robot being uncertain about a particular word is the presence of other words in its memory that were said in exactly the same situation. If this is the case, the robot might be undecided, which of these words it should say, and therefore hesitates. After having said a word, and if the situation does not change, it will look for another word in its memory that also matches the given situation. If the situation does change, it may say the same word again, that it said just before, or others, depending on whichever word fits best to the corresponding situation. The robot stops speaking as soon as the box falls down or if participants put it back on the table.

As we will later see in the analysis in section 5, the three biggest shortcomings of this system is the circumstance, that the robot does not know, when it is being addressed by the participant, that it does not know what is being said at the time of the interaction, and that it neither knows how, what is said, is being said, in terms of pitch, energy etc. In terms of the importance of these shortcomings, we consider the first one (not knowing that it is being spoken to) as the most fundamental one, followed by the last one (not knowing how it is being spoken to).

3.5 Body Memory

The *body memory* saves all high- and low-level perceptions as well as behaviour ids and motivation values as they occur in a file and timestamps them. Only auditory perceptions are excluded from this module and processed separately by the *auditory system*. The resulting sensorimotor-motivational file is subsequently used by the *lexical grounding system*.

3.6 Auditory System

The auditory system comprises all processes related to the extraction of lexical units, which are words in the experiments reported upon within this thesis. The auditory system currently works offline in between experimental session. The following processes may be distinguished:

- speech recognition and word alignment
- prosodic labeling
- word extraction

3.6.1 Speech recognition and word alignment

In previous experiments on the acquisition of object words described in (Saunders et al. 2010) it was found that regular off-the-shelf speech recognition software did not yield a sufficient accuracy when applied to the speech used by their participants within the human-robot interaction. As the experimental scenarios described in chapter 4 are very similar to the scenarios described in (Saunders et al. 2010) we utilized the speech recognition and alignment procedure outlined there. This means that instead of using a fully-automated but not very accurate standard speech recognizer, a combination of two half-automated systems was employed which rely on manual transcription of the speech recording and a subsequent manual re-alignment of word timings and audio stream. We will just give a short sketch of the system and refer to (Saunders, Lehmann, Sato and Nehaniv 2011) for details.

The first part of the system, basically a specialised speech recognition system, takes a cleaned audio-file, that contains the recorded speech, and a separate transcript of this very speech recording as input and outputs a timed phonetic transcription. This file is

further processed and subsequently, together with the original audio recording, loaded into the second sub-system, which performs a prosodic analysis. Before this analysis can be performed, the phonetic transcription has to be manually re-aligned in order to improve the precise timing of the words relative to the audio recording.

The time effort of the two manual steps is considerable and possibly only surpassed by a full-fledged conversation analytical transcription. On average 5 minutes of audio recording, the length of one experimental session, required 4 hours of post-processing.

3.6.2 Prosodic Labeling

In order to be able to compare the effect of introducing affect into the language acquisition architecture as opposed to not having any affective values, the same prosodic labeling system was employed as in (Saunders, Lehmann, Sato and Nehaniv 2011). This system takes as input the correctly-timed transcription file, i.e. the output from speech recognition and word alignment, and produces for each utterance a set of prosodically salient words which is subsequently used by the word extraction mechanism.

3.6.3 Word Extraction

In initial experiments a set of heuristics was tested, which was developed by (Saunders et al. 2010) and which is not related to prosody: there, a word is considered salient if it is firstly utterance-final, with the end of an utterance being defined as a pause longer than average word duration, and if, secondly, its duration is longer than average.

In the case of participant *P04*, who was the first real participant of our experiments, the just described word extraction mechanism was employed. For all other participants the following one is used: from each utterance that word is extracted which has the highest prosodic salience as calculated by the prosodic labeling system. Upon the execution of a full

set of test-sessions with participant *P04*, we decided to use the latter mechanism as none of the two mechanisms appeared to be superior. The said mechanism has the additional advantage of being the word extraction mechanism employed within the experiments with the "affect-less" architecture of (Saunders, Lehmann, Sato and Nehaniv 2011). This means that choosing the same word extraction mechanism as (Saunders, Lehmann, Sato and Nehaniv 2011) renders their and our results comparable.

The application of the three mechanisms above to a speech recording from an experimental session results in a file that contains only salient words and the time at which they were uttered relative to a start time stamp (cf. figure 3.7, top-left `*.emph_words` file). Furthermore, the start and end time of a particular word are extended to the start and end time of the entire utterance. One could say that the most salient word of an utterance gets to represent the whole of the utterance (see also Saunders et al. 2010).

3.7 Lexical Grounding System

The lexical grounding system performs, as the name indicates, the grounding of the lexical items produced by the auditory system. This process is performed offline, after the execution of an experimental session. It takes as input both the sensorimotor-motivational file as recorded the body memory during that session, and the salient word file which is generated by the auditory system after the session. Both files are subsequently merged into one file in which each salient word is associated with the *smm* data that was recorded at the time during which the corresponding utterance was produced. In the current implementation we made the decision to eliminate duplicate entries, that originate from the same utterance, in order to keep the lexicon at a manageable size. The grounding process is depicted in figure 3.7.

The resulting file, in figure 3.7 `P05-011211-british.laction`, is merged with the embodied lexicon from previous sessions, if any, to form an updated embodied lexicon which is subsequently used in the followup session by the languaging system.

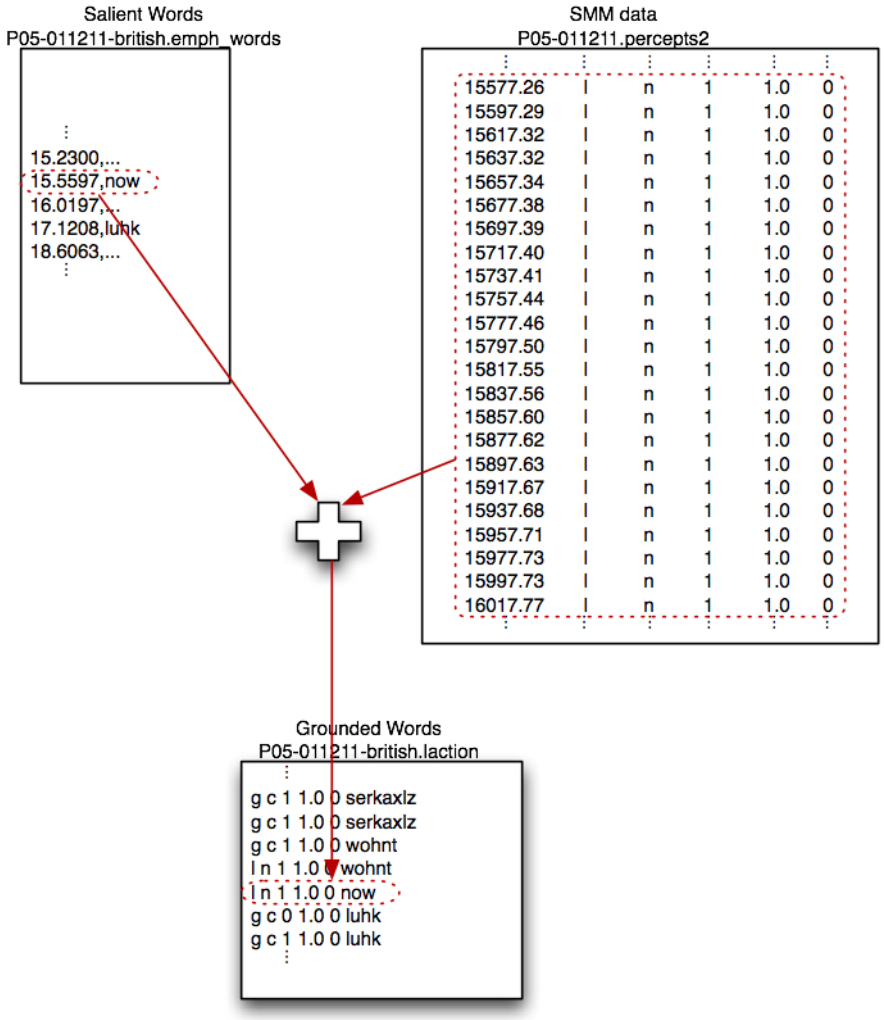


Figure 3.7: ***Grounding of (salient) words****. The grounding process associates lexical entries, in our case salient words, with the concurrently occurring sensorimotor-motivational data. In our system the salient word is propagated across the entire duration of the utterance, such that the time stamps, visible in the salient-words-file (top-left) do mark the start and end of the respective utterance within which the word was produced. Time stamps for utterance boundaries are symbolized by '...'. Also notice that we remove duplicates of grounded words that would ensue from the same utterance. In the given example this means that due to the lack of change within the* smm *data while the utterance was produced the potentially ensuing 23 identical grounded words are collapsed into one (bottom).*

Die Grammatik des Wortes "wissen" ist offenbar eng verwandt der Grammatik der Worte "können", "imstande sein". Aber auch eng verwandt der des Wortes "verstehen". (Eine Technik 'beherrschen'.)

—*Ludwig Wittgenstein, PU 150*

Chapter 4

Experiments on the Acquisition of Negation

4.1 Introduction

In order to test the two hypotheses outlined in section 1.1, we devised two experimental scenarios, a so-called *rejective* scenario and a so-called *prohibitive* scenario[1], and embedded them in two experiments, which were subsequently called the *rejection* and the *prohibition experiment*. Both scenarios target the acquisition of the earliest types of negation such as the rejective "no", and both scenarios were grafted on top of the experimental scenario developed by (Saunders et al. 2010) and (Saunders, Nehaniv and Lyon 2011)[2].

Differing from Saunders' experiment both the *rejection* as well as the *prohibition* experiment are what psychologists refer to as *blind experiments*. Within Saunders' experiment, participants were told to teach the humanoid "Deechee" about particular objects, i.e. the

[1] We will alternatively refer to these scenarios as *rejection* and *prohibition scenario*.

[2] We will in the following for reasons of readabilty refer to these scenarios and experiments as *Saunders' scenario* and *Saunders' experiment* respectively. By doing so we do not intend to deny the contribution of the other authors.

objects' names, their colour, and their size, and this was indeed what Saunders' experiment was about. By contrast, we largely assumed those teaching instructions, yet, in our case, they served as a cover story. This means that our participants did not know about the true purpose of the experiments. At no time did we tell them that, what we refer to in here as *negation experiments*, were about the acquisition of *negation*.

This choice was made, as it was unclear to the author, how one could potentially go about to teach *negation*. *No* and words with similar functions in other languages, seem to be so embedded within the fabric of language, that we had no intuition, what one could and would possibly do, to teach 'the meaning' of *no*. Thus, as we could not imagine what participants would do, if we told them to teach the robot 'negation', in combination with potentially willful artificial modifications of participants' speech during such unimaginable doings, we decided not to tell them at all. Instead, we decided to elicit negative words naturally by putting participants in situations similar to those in which, according to the developmental literature, negative words naturally occur and from which they are thought to originate (cf. section 2.5).

Yet, despite the tendency of caregivers to interpret a toddler's emotional state linguistically via so called *intent interpretations*, we could not be sure, if participant's would do the same for our robot. A side-effect of the circumstance that they did, as we will show in chapter 5, is that this result strengthens hypothesis 1.

4.2 Instructions to participants

Thus, in both the rejection as the prohibition experiment the naïve participants were given very similar instructions as in Saunders' experiment, namely that their task was to teach Deechee about the available objects. The objects consisted of boxes with different shapes

printed on them such as star shapes, heart shapes, circle shapes etc, which constitute tags from the ARToolKit package as described in chapter 3. Moreover, participants were told to imagine Deechee to be a small child of approximately 2 years and, further, that Deechee had preferences for particular objects, that it may like, dislike, or would feel neutral about. The first additional instruction, to imagine Deechee to be a 2-year-old child, was given in order to increase the likelihood of participants assuming a simplyfied speech register akin to *CDS*. The second additional instruction about Deechee having preferences, was given in order to prepare participants for Deechee's emotional displays. As opposed to a child, humans may not expect a robot to display emotions or intentional behaviour. Thus, in order to render the experimental situation more similar to that of a mother-child dyad, we prepared them for emotional displays.

All objects were situated on a table between the participant and Deechee. The participants were further instructed that they should try to teach Deechee about all of the boxes even though Deechee might not like particular ones, and that Deechee would communicate its preferences to them. No further explanations were given as to how the humanoid communicates its preferences. Participants were told that they could put the boxes into the humanoids hand if they wanted to. It was also mentioned to the participants that Deechee would not know a single word at the start of their set of sessions, and that it might start to speak from session 2 onwards. If participants asked for more instructions as to how exactly to speak to Deechee, it was only emphasized that they should speak to it as if it was a small child and that they should say whatever they would deem appropriate in the given situation. It was further emphasized that we could not give them more precise instructions in terms of their language use.

The written instructions and consent forms are available on the appended DVD (cf. section D).

4.3 Recruiting and distribution of participants

We recruited 10 participants per experiment, all but one of which were native English speakers, the majority British, with one Asian (the only non-native), one US-American, one Nigerian, and one South-African speaker, the latter of which had lived for several decades in the UK. The participants were balanced by gender. Further balancing was not possible, due to the difficulty in recruiting participants for an experiment for which they had to return to the lab five times. Most of the participants were recruited from the campus and were either students or employees, only three of them had no affiliation with the university. Participants were remunerated with £20 each after completion of all five sessions.

4.4 General experimental setup

Each participant completed five sessions of approximately five minutes each with at least one day in between sessions, which was needed for the post-processing of the speech recordings (cf. chapter 3). All participants were wearing headsets during the interaction in order to record their speech and we further videotaped each session. Apart from the participant, one to two more people were in the room: an operator, who started up the system and monitored it during the session, and a helper which placed the boxes back on the table, as Deechee was prone to drop them. In a few sessions, the helper was absent and the operator took on both roles. Participants were seated opposite the robot with a table separating the two. The five objects, which were to be taught, were cardboard boxes of approximately 10cm side length, with black-and-white symbols printed on each side of each box. For any particular box, the symbols were identical on every side. The symbols were a star, a heart, a square, a crescent moon, and a triangle. After having read and signed the instructions

and the consent form, participants were seated in room and asked to count down in the following way: "three, two, one, start". They were told in advance, that "start" would constitute the start time of the experiment and that the operator would signal when the five minutes were over. Upon "start", the operator pressed a button, which subsequently produced a time stamp within the body memory.

After completion of the fifth session the experimenter carried out a short interview with the participants, in which they were asked to evaluate the learning progress of Deechee, and about particular aspects of its linguistic and non-linguistic behaviour including questions relating to negation. During this interview they could give general comments on the experiment if they wanted to. The entire interview was recorded and transcribed. Moreover, each participant was asked to fill out a *Ten Item Personality Measure (TIPI)* form (Gosling et al. 2003) in order to detect potential correlations between personality traits and linguistic behaviour. Neither interview nor *TIPI* form have been evaluated at the time of writing due to time constraints associated with the PhD programme.

4.5 Rejective Scenario

This scenario was devised in order to test if intent interpretations might be sufficient for the acquisition of early types of negation such as rejective utterances or motivation-dependent denial. An example for rejection would be a simple "no" as in "no, I don't want potato mash" when being handed some. An example for motivation-dependent denial would be another simple "no", but as an answer to a motivation-related question such as "Do you want some potato mash?" (see coding scheme in section B.8). Caretakers of young children were observed to engage in intent interpretations, and these are hypothesised to play an important role in the acquisition and eventual linguistic expressions of intention

(cf. hypothesis 1 in section 1.1).

Imagine a child that does not speak yet and which rejects a particular object or action by "turning her head aside, pushing or throwing the aversive thing away" etc. Parents "frequently interpret these behaviors as expressive of negation and expand them with lexical negatives: 'no, no, don't want it' " (Pea 1980).
The rejective scenario was designed to elicit intent interpretations of this type from participants by having the humanoid express non-linguistic forms of rejection similar to the aforementioned ones: Deechee, upon registering the presentation of a disliked object, will briefly look at it, start to frown and turn its head away.

4.5.1 Parameter settings

Table 4.1 shows the object-bound motivation values as they were set for each session. They are identical for each participant in order to make the various sessions comparable across participants. Note that during the entire 5 sessions each object triggers negative and positive motivation twice, and neutral motivation once.

Time constants, that are important to maintain the interaction, were chosen as depicted in table 4.2. The particular choices were determined through fine-tuning during the design phase with the author's intuition as sole evaluation criterion.

	session 1	session 2	session 3	session 4	session 5
triangle	1	-1	1	-1	0
moon	0	1	-1	1	-1
square	-1	0	1	-1	1
heart	1	-1	0	1	-1
circle	-1	1	-1	0	1

Table 4.1: *Object-bound motivation values per session for both scenarios.*

variable	*value*	*description*
face_time	0.8	duration of iCub looking at face when pickup detected and motivation ≥ 0
object_time	3	duration of iCub looking at object when *pickup* detected and motivation ≥ 0
dwell_time_face	1.2	duration of iCub looking at face when no *pickup* detected
dwell_time_object	2	duration of iCub looking at obj. when no *pickup* detected
maxIdleTime	3	perceptual timeout for high level percepts: if no objects or faces are perceived for *maxIdleTime*, iCub looks back at the table

Table 4.2: *Important constants for human-robot interaction, all values in seconds*

4.6 Prohibitive Scenario

The prohibitive scenario takes the rejective scenario as a basis but extends it to include the elicitation of prohibitive utterances from the participants. In this scenario, two of the boxes were declared to be *forbidden*, and were marked with coloured dots on the side facing the participants. The latter were told, in addition to what they were told in the rejective scenario, that the marked objects were *forbidden* objects, and that Deechee was not allowed to touch them. In order to keep Deechee from touching these forbidden boxes, participants were instructed to physically restrain the robot, in case it tried to approach them. Thus, before the first session, participants were shown how to push the robot's arm back, firstly, in order to show them the ideal contact point, such that the robot's hand would not be damaged, and secondly, to take their potential fear from actually touching the robot. The ideal contact point is the wrist and lower arm. In this scenario force control was used as the control mode for both arms, which makes it possible for participants to manipulate the arms while the robot executes a movement. The act of pushing the robot's

arm is detected as physical resistance and registered by the perception system as *resistance event*. The system broadcasts this information to the other systems, such that the body memory registers and saves this event together with the other percepts (cf. section 3 for more details on the technicalities). The occurance of such a resistance event further leads to the motivation being set to negative: Deechee subsequently starts to frown. In this case Deechee will also look slightly longer at the participant's face (see parameters below). We decided to elongate the gaze durations, when grumpy, in order to make sure that participants would actually notice Deechee's grumpy despair, and further, to give this emotion a slightly higher intensity. The elongation of a gaze time by 0.8 seconds may not seem much to the conversationally untrained reader, but the human conversational system is highly sensitive to differences in pause durations which typically range within the 0.1 - 1 second range, with 1 second being hypothesized to be a significant threshold (cf. section 2.1.3). We did not retrieve exact information for standard gaze times, but due to the tight integration of gaze with speech in human conversation, we were fairly confident, that a difference of 0.8 seconds would not go unnoticed - if not consciously, then subconsciously.

This scenario was designed to test hypothesis 2, which relates to the origin of linguistic negation (cf. section 1.1). Under this hypothesis one assumes that the very root of linguistic rejection, typically the first negation type to emerge, originates from the prohibitive use of "no" by caretakers, and that it is produced in conjunction with physical restraints. The latter are typically applied in order to keep the child from doing whatever it intended to do, for example putting its finger into an electric socket.

Physical restraint is not the only way in which a child can be prohibited from doing something. For example, smaller dangerous or otherwise forbidden objects can be put out of reach of a child. We would have preferred not to give the participants any particular instructions as to how to prohibit Deechee from touching a *forbidden* object. But due to

the limiations of the robot's vision system, this was not a viable option. To our displeasure, many participants did not follow our instructions, despite of the initial "contact exercise" (cf. section 5.5), and did not physically restrain the arm movement. Yet the ensuing interaction for those who did looked very promising and fulfilled our expectations.

4.6.1 Parameter Settings

In order to keep the results of prohibitive and rejective scenario as comparable as possible, the object-bound motivation values were identical to those in the *Rejective scenario* (see table 4.1). Also the time constants that largely determine the robots interactive behaviour are the same as in the *rejective scenario* (see table 4.2). Additionally the following two time constants were introduced to adapt the robot's behaviour in case of physical restraint (see table 4.3).

variable	*value*	*description*
grumpy_face_time	1.6	duration of iCub looking at face if physical restraint is detected (this implies that iCub was reaching for an object)
grumpy_object_time	2	duration of iCub looking at object if physical restraint is detected

Table 4.3: *Additional time constants for human-robot interaction in prohibitive scenario, all values in seconds*

4.7 Study Design: Comparison of Hypotheses

Both of the abovementioned scenarios were either embedded or singularly constitutive for the two experiments and arranged such that the impact of the so called prohibition task

could be directly compared between the two experiments. Upon the advice of a psychologist within our research group, we decided to arrange the experiments as indicated in figure 4.1. The rejection experiment consists solely of five sessions of the rejection scenario and was

		Prohibition Experiment	**Rejection Experiment**
		Group 1 (Test Group)	**Group 2 (Control Group)**
sessions 1-3		*Prohibition + Rejection*	*Rejection only*
	Robot	{L D D L N}	{L D D L N}
	Human	{A A P P A}	{A A A A A}
sessions 4+5		*Rejection only*	*Rejection only*
	Robot	{L D D L N}	{L D D L N}
	Human	{A A A A A}	{A A A A A}

Figure 4.1: ***Study Design****: {..}: Permutations of the given values;* L*: Liked object,* D*: Disliked object,* N*: Neutral object,* A*: Allowed Object,* P*: Prohibited/Forbidden object. The mapping of* positive/negative/neutral *valences to objects was permutated between sessions, such that each object was twice* liked*, twice* disliked*, and once* neutral *across the 5 sessions. The mappings were identical for every participant. The* allowed/forbidden *markers were permutated as well (see text).*

executed first. From the very start we could see, that our participants did indeed engage in the envisioned *negative intent interpretations*, that are at the core of hypothesis 1 (cf. section 1.1). Subsequently this scenario was chosen to constitute the base line against which the prohibition scenario, which is linked to hypothesis 2, could be compared. Upon said psychological advice the prohibition experiment was designed as composite experiment, where the first three sessions contain the treatment, here the presence of the prohibitive task, which together with the base line task forms the prohibition scenario. The prohibitive

task was then removed from the last two sessions of the prohibition experiment, in order to determine the effect of the treatment upon the acquisition of negation by the robot. The assessment of the two hypotheses is then based on the comparison of the robot's (negation) performance within the treated group with its performance within the baseline group during the last 2 sessions.

The mapping of negative, positive, and neutral valence to the five available objects was permutated in each session such that each object had twice a positive, twice a negative, and twice a neutral valence asisgned to it. This was done for several reasons: Firstly, we wanted to prevent participants from knowing from a previous session what Deechee would like within the current session as they might have totally avoided or engaged less in talking about the unliked objects. This naturally would have decreased the frequency of potential negative intent interpretations as the latter mainly occur with negative emotional displays. Secondly, we intended to prevent the robot learning the association between negative words and certain objects. The latter could happen if these objects had constantly a negative valence assigned to them for, as far as the learning algorithm goes, this would constitute an attribute of the object rather than an attribute of the robot.

For the prohibition scenario the assignment of the *forbidden* and *allowed* attributes, which only participants were aware of, was done in such way that every combination of *liked*/*disliked* with *allowed*/*forbidden* would occur at least once within each session. This, together with the change of the valence-to-object mapping between subsequent sessions then lead automatically to a permutation of the *allowed* and *disallowed* attribute-to-object mappings across sessions. In general there were either two or three *forbidden* objects per session, and two or three *allowed* ones.

A note on practicalities As a last remark within this chapter, we would like to give some realistic estimation of the time effort involved in these experiments for those readers which may consider untertaking a similar endeavor.

The rejection experiment was performed between October 2011 and May 2012. Overlapping the former, the prohibition experiment was performed between February and April 2012. It thus took us about 7 months to complete the 100 sessions of both experiments. The most important reason for this long duration is the enormous effort that comes with the semi-automatic speech processing (cf. section 3.6). The post-processing of one session of 5 minutes required on average about 4 hours. This means that the number of man hours, generated by post-processing alone, amounted to 400 hours or 2 - 3 man-months. Luckily we received help from a lab assistant such that the time effort could be distributed. The second important reason for time-delays is the circumstance that participants have to return four times to the laboratory, which leads to delays, due to the lack of availability of the latter during longer stretches of time. Other factors that contribute to delays in the execution of the experiment have to do with failure or breakage of the robot, though the latter is, surprisingly to the roboticist, a minor factor compared to the former two.

It is therefore our contention that an improvement to the semi-automatic speech processing should be considered if more experiments of this kind should be performed in the future. We will pick up this topic in the discussion section in chapter 6.

P08 ((presents D with moon box, who looks at it and smiles))
P08 Do you like the moon? ((D approaches box, flinches back))
(2.0)
D No
P08 I think you do ((P holds box further out towards D))
(0.8)
D go
P08 ((chuckles)) .hh allright I'm not gonna force [you
D [no] (2.1) go
P08 When a robot says no I should presumably just (.) relax

—Participant 08 and Deechee, session 2

Chapter 5

Making Sense of Negative Utterances

This chapter constitutes the core of this thesis and might have been simply titled "Analysis". As will be explained below, different variants of analysis will be employed which complement each other to a certain degree. Before going into detail some frequently occurring words shall be clarified.

5.0.1 Negation words, negative utterances, and negation types

Negation words/Negative words

The expressions *negation word* and *negative word* are used in the following interchangeably. The negation words listed in table 5.1 form the basis of the subsequent analyses and were selected manually from a list of words which is derived from the entirety of transcripts gathered during the experiments. All of these words are either lexically negative, such as *no*, or they are part of grammatically negative constructions, such as *not* or *don't*. Two negation words have been excluded from the analysis due to each of them occurring exactly once in the whole corpus: *aren't* and *nah*. The latter might not be standard in written English but was frequently found to substitute the more regular *no* in spoken English.

The only non-standard negation word that was observed and which made it into the list is *nono*. *Nono* was used by one participant in a noun-like manner such as in "this is a nono". The participant produced this utterance in order to indicate to Deechee that a box was off-limits.

Some lexical negation words have more than one phonetic equivalent. This might be due to the coders either choosing a different phonetic transcription of a word, but it also might be caused by the participant speaking in a different accent. The lexical *don't* for example was found to have the two phonetic counterparts *duhnt* and *downt*. As our system processes only phonetic words, these count as different words to the robot. We therefore treat them as separate lexical units and add a '2' in brackets to distinguish them in their non-phonetic, standard English notation.

Table 5.1: ***List of negation words used for analyses****. All negation words listed here were selected from a complete list of words obtained by accumulating the words from the transcripts of the experiments. A trailing '(2)' signifies a second phonetic variant of the same lexical word.*

no	don't	don't (2)	not	didn't	didn't (2)
isn't	won't	can't	can't (2)	wouldn't	doesn't
doesn't (2)	couldn't	wasn't	weren't	haven't	hasn't
mustn't	cannot	shouldn't	nono	neither	

Negative utterances

In the following an utterance is considered a negative utterance if it contains at least one negation word. The majority of negative utterances contain exactly one negation word, but utterances with two negation words are not uncommon either. As the utterances of our participants were extracted automatically, or, in other words, the utterance boundaries were detected automatically, the ensuing utterances may not coincide exactly with what a

human transcriber would conceive of as an utterance. But roughly speaking, and assuming an optimal utterance isolation mechanism the notion of *utterance* coincides with the intuitive notion: a singular move in speech that is either separated from a subsequent move by a small gap or that is terminated by the move of another speaker. One may think of utterances as atomic or smallest complete units in spoken language. Yet it is the author's conviction that some of these units are not so complete after all, especially if a less-than-perfectly competent speaker is involved. Parents often "auto-complete" their children's' utterances, and the literature reports the construction of utterances across speaker turns by mother-child dyads. Our intuitive notion of an utterance, i.e. the conversational move of a single speaker, as being "complete" therefore might be overly simplistic. Yet for the time being, the reader may stick to this notion and use it as a mental crutch. Numerically speaking the relationship between negative utterances and negative words is a $1 : n$ relationship.

Examples:

[P05,s5] *You don't like the heart?* (negative utterance with 1 negation word)

[P15, s2]*No, you don't like that one.* (negative utterance with 2 negation words)

Negation types

The notion of negation types will be introduced more thoroughly in section 5.3. They distinguish negative utterances on the pragmatic level. Roughly speaking they are a distinction of negative utterances into different types of speech acts, if the notion of *speech act* is conceived in a liberal fashion. Producing a negative utterance implies that an agent engages in at least one type of negation. On rare occasions the production of a negative utterance during the experiments was classified as an agent engaging in more than one

type of negation. Thus formally the relationship between negative utterances and negation types is a $1 : n$ relationship, yet in the overwhelming majority of cases the relationship is $1 : 1$.

Examples:

[P15, s2] *You're not allowed to play with this one, I told you you're not allowed to play with this one Deechee.* (1 negative utterance with 2 negation types: *prohibition* followed by *apostrophised negation*)

[P11, s3] *You don't like moon, do you? No.* (1 negative utterance with 2 negation types: *neg. mot. question* followed by *neg. intent interpretation*)

5.0.2 Analytical Methods

We will in the following use three different methods to look at the participants' as well as the robot's negative utterances. There are two essential reasons for employing more than one method. Firstly, these three methods focus on different levels of negative utterances, with a slightly different granularity on each level.

On the utterance level (section 5.1) we will look at negative utterances as a whole in terms of rather conservative linguistic measures such as mean utterance length (MLU), the number of distinct words, and the number of utterances per minute. Some of these measures can be used to infer cognitive expectations and ascriptions on part of the participants towards the robot. Furthermore the MLU could, at least principally, be compared to the MLU of child-directed speech to determine, if similar adaptations can be observed. Moreover this analytical level allows us to conduct statistical comparisons between the two negation experiments and Saunders' experiment, in order to determine if one of our main hypothesis holds: Does the display of emotions and motivationally congruent behaviour

indeed lead to a form of speech on part of participants that involves negation.

On the word level (section 5.2) we will aggregate the speech from all participants and sessions into single, big corpora - one per experiment. There we will look at the absolute and relative frequencies of word and word groups and compare the two corpora in between each other as well as to other corpora in order to detect any impact of our experimental setup on the word level.

On the pragmatic level (section 5.3) we will introduce and evaluate a pragmatically and conversation analytically driven taxonomy of negation types and apply this taxonomy to the participants' negative utterances. On this level we will see most clearly the impact of the two different settings upon the speech of our participants. It is our contention that this analytical level is the explanatorily most valuable of the three levels as it is only here that we see what participants were actually doing when producing their negative utterances. Moreover, we will link the pragmatic with the word level in this section and have a closer look at our notion of prosodic saliency.

In section 5.4 we will evaluate the robot's negative utterances on a pragmatic level and in terms of the adequacy of these utterances in context. It is also here where we will pitch the two hypotheses on the origin of negation against each other as outlined in the description of the experiments.

In section 5.5 we will take a closer look at the temporal coordination between and correlation of prohibitions (one of our negation types), prohibitive corporal actions, and the motivational state of the robot. This section was necessitated by a, to us, surprising result in the preceding section.

The other reason for having several analytical levels, especially in language-related research, is the circumstance that some academic groups may disregard the pragmatic level, especially as it is based on the subjective assessment of a coder. As the reader

will see, the impact of the robot's motivated behaviour can be shown on each analytical level. That is, even if one fosters absolute disregard for all things pragmatic, the 'woolly', subjective results on the pragmatic level are mirrored on the 'lower' levels in terms of hard word counts and frequencies and in the number of negative utterances.

5.1 Human: Utterance Level

In this section an analysis of the speech of our participants on the utterance level is performed using basic measures that can be calculated automatically once a transcription of the participants' speech, segmented into utterances, is available. These measures characterize the participants' linguistic behaviour on a rather coarse-grained level in terms of verbosity, "how much do participants speak", and utterance complexity (cf. Fischer et al. 2011). Said measures are calculated for the complete speech data of all participants, separated by experiment, but also for negative utterances only. The tables 5.2 and 5.3 give an overview of these measures for all utterances in the rejection and prohibition experiment respectively.

5.1.1 Measures on the utterance level

mean length of utterance (*MLU*) This is one of the most common measures in the literature on child language development (cf. Owens 2012[1]). It is used on the one hand to estimate the linguistic capabilities of a child in terms of utterance complexity. On the other hand it is also used to show how parents and caretakers simplify their speech in terms of utterance length when speaking to children. We use the word-based MLU[2] which is obtained by dividing the number of phonetic words by the number of utterances per speaker and session (cf. Fischer et al. 2011, Roy et al. 2009).

[1] Owens (2012) discusses in chapter 2 some of the limitations of this measure in terms of its explanatory value and emphasizes that the child-based MLU might vary considerably across situational contexts. If this should also hold true for the caregivers MLU, this could have important repercussions for the comparative value of MLU across experiments in HRI. HRI experiments with a linguistic focus are to date extremely limited and specific in their situational context.

[2] Alternatively the MLU can be calculated based on morphemes. This means that the MLU used with this thesis is not necessarily comparable to the morpheme-based MLUs.

Table 5.2: ***Utterance-level Measures for Rejection Experiment****. All participants and all sessions. Any given number refers to the participant with participant id noted on top the corresponding column and the session number in the corresponding first column. Abbreviations: sX: session nr. X, # w/# u: total number of words/utterances uttered by participant, # dw: number of distinct words, MLU: mean length of utterance, w/min / u/min: average number of words / utterances per minute*

		P01	P04	P05	P06	P07	P08	P09	P10	P11	P12
s1	d (s)	272	168.4	308.6	178.3	380.9	290.4	301.9	275.5	298.3	300.5
	# w	111	314	370	392	729	505	474	26	458	353
	# u	50	67	133	104	225	157	126	16	114	92
	# dw	21	101	96	100	140	145	133	6	103	114
	MLU	2.2	4.7	2.8	3.8	3.2	3.2	3.8	1.6	4	3.8
	w/min	24.5	111.9	71.9	131.9	114.8	104.3	94.2	5.7	92.1	70.5
	u/min	11	23.9	25.9	35	35.4	32.4	25	3.5	22.9	18.4
s2	d (s)	285.4	196	305.2	293.2	303.4	259.8	306.6	287.6	298.8	312.7
	# w	138	346	323	580	666	414	437	70	338	230
	# u	66	83	110	156	183	112	122	36	106	82
	# dw	26	77	71	86	107	89	138	27	90	76
	MLU	2.1	4.2	2.9	3.7	3.6	3.7	3.6	1.9	3.2	2.8
	w/min	29	105.9	63.5	118.7	131.7	95.6	85.5	14.6	67.9	44.1
	u/min	13.9	25.4	21.6	31.9	36.2	25.9	23.9	7.5	21.3	15.7
s3	d (s)	307.9	297.1	249.9	318	307.8	296.8	299.2	290.8	306.6	302.5
	# w	159	468	302	662	569	431	513	62	242	107
	# u	66	111	102	168	180	128	139	31	98	38
	# dw	29	107	73	97	100	96	155	27	56	22
	MLU	2.4	4.2	3	3.9	3.1	3.4	3.7	2	2.5	2.8
	w/min	31	94.5	72.5	124.9	110.3	87.1	102.9	12.8	47.4	21.2
	u/min	12.9	22.4	24.5	31.7	35.1	25.9	27.9	6.4	19.2	7.5
s4	d (s)	319.2	265.5	213.2	329.9	307.8	301.5	300.4	289.4	300.2	303.5
	# w	204	393	253	685	541	364	495	84	187	152
	# u	80	93	89	187	184	105	143	33	70	67
	# dw	30	104	67	95	108	88	152	23	58	27
	MLU	2.5	4.2	2.8	3.7	2.9	3.5	3.5	2.5	2.7	2.3
	w/min	38.3	88.8	71.2	124.6	105.5	72.4	99.1	17.4	37.4	30.1
	u/min	15	21	25	34	35.9	20.9	28.6	6.8	14	13.2
s5	d (s)	305.9	220.4	269.1	307.1	314.5	324.5	301.3	290.2	319.5	299
	# w	160	370	356	633	582	406	402	66	211	134
	# u	61	92	135	157	188	128	132	36	74	54
	# dw	23	104	81	89	105	127	116	20	74	36
	MLU	2.6	4	2.6	4	3.1	3.2	3	1.8	2.9	2.5
	w/min	31.4	100.7	79.4	123.7	111	75.1	80.1	13.6	39.6	26.9
	u/min	12	25	30.1	30.7	35.9	23.7	26.3	7.4	13.9	10.8

Table 5.3: ***Utterance-level Measures for prohibition experiment.*** *All participants and all sessions. Any given number refers to the participant with participant id noted on top the corresponding column and the session number in the corresponding first column. Abbreviations: sX: session nr. X, # w/# u: total number of words/utterances uttered by participant, # dw: number of distinct words, MLU: mean length of utterance, w/min / u/min: words/utterances per minute*

		P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
s1	d (s)	301.8	317.8	311.8	332.3	303.8	301.1	318.9	300.4	359.3	319.3
	# w	535	279	611	704	200	691	644	508	332	404
	# u	134	92	165	194	66	185	179	138	119	124
	# dw	110	76	134	195	63	170	178	114	74	99
	MLU	4	3	3.7	3.6	3	3.7	3.6	3.7	2.8	3.3
	w/min	106.4	52.7	117.6	127.1	39.5	137.7	121.2	101.5	55.4	75.9
	u/min	26.6	17.4	31.8	35	13	36.9	33.7	27.6	19.9	23.3
s2	d (s)	324.2	305.7	301.4	307	310.7	296	308.4	308.3	317.6	312.1
	# w	653	307	715	830	215	702	535	610	363	332
	# u	178	93	187	204	78	188	147	138	133	110
	# dw	100	73	126	187	53	188	117	110	75	77
	MLU	3.7	3.3	3.8	4.1	2.8	3.7	3.6	4.4	2.7	3
	w/min	120.9	60.3	142.3	162.2	41.5	142.3	104.1	118.7	68.6	63.8
	u/min	32.9	18.3	37.2	39.9	15.1	38.1	28.6	26.9	25.1	21.1
s3	d (s)	297.4	332.5	294.6	326.8	302.3	309.1	316.2	302.6	306.1	315.8
	# w	424	343	717	774	184	702	610	625	329	477
	# u	133	95	180	221	71	195	170	137	141	162
	# dw	70	70	152	205	52	178	130	87	82	100
	MLU	3.2	3.6	4	3.5	2.6	3.6	3.6	4.6	2.3	2.9
	w/min	85.5	61.9	146	142.1	36.5	136.3	115.8	123.9	64.5	90.6
	u/min	26.8	17.1	36.7	40.6	14.1	37.9	32.3	27.2	27.6	30.8
s4	d (s)	307.6	319.9	308.5	316.1	314.7	314.8	301.7	298.6	301.1	316.2
	# w	501	298	698	714	259	750	536	490	332	589
	# u	154	89	181	192	83	198	160	131	119	204
	# dw	69	55	132	195	40	194	117	93	59	127
	MLU	3.3	3.3	3.9	3.7	3.1	3.8	3.4	3.7	2.8	2.9
	w/min	97.7	55.9	135.7	135.5	49.4	143	106.6	98.5	66.2	111.7
	u/min	30	16.7	35.2	36.4	15.8	37.7	31.8	26.3	23.7	38.7
s5	d (s)	306.7	380.3	312.4	320.1	293.4	302.1	311.2	303.2	306.6	317.4
	# w	476	340	656	728	310	716	591	577	333	493
	# u	160	100	186	188	102	199	178	157	127	170
	# dw	64	52	131	198	42	176	109	105	69	106
	MLU	3	3.4	3.5	3.9	3	3.6	3.3	3.7	2.6	2.9
	w/min	93.1	53.6	126	136.5	63.4	142.2	113.9	114.2	65.2	93.2
	u/min	31.3	15.8	35.7	35.2	20.9	39.5	34.3	31.1	24.9	32.1

It is important to note that MLUs are not comparable across languages (Owens 2012). Fischer et al. (2011) counts the MLU amongst the measures for linguistic complexity. Self-evidently this measure is sensitive to differences in the segmentation of speech as the latter impacts utterance length. As an automatic system for speech segmentation was used in our experiments (cf. section 3.6), the MLUs reported here can currently not be compared with MLUs from HRI studies that use different systems or MLUs obtained from speech that was manually segmented.

number of utterances per minute (*u/min*) The number of utterances per minute reveal how much participants spoke to the robot, if they were talkative or taciturn - a measure of linguistic verbosity. To account for the variability of session length, this measure will be used instead of the total number of utterances.

number of distinct words (*#dw*) This measure indicates if participants used a variety of words when speaking to the robot or if they were repetitive instead. Fischer et al. (2011) classes this measure, together with MLU, amongst the measures for linguistic verbosity.

5.1.2 Potential impact of measures upon the language acquisition system

Mean length of utterances

Firstly, the length of an utterance (*UL*) is relevant to the language acquisition of the robot in that precisely one word of each utterance, namely the most salient one, is extracted and subsequently becomes part of the robot's active vocabulary (see section 3.6). Thus, the longer the MLU, the less words will be propagated into the lexicon and the bigger the statistical impact of each word.

Secondly, Saunders, Lehmann, Sato and Nehaniv (2011) report a 71% accuracy of the prosodic labeling system with regards to the extraction of the most salient word. They do not specify if there is a correlation between error rate and utterance length, but intuitively this seems likely. Consider for example the following two utterances (prosodically emphasized words, i.e. words with high energy and higher pitch are capitalised): "It was the CAT" (UL=4), "It was the CAT, NOT the dog"[3] (UL=7). In longer utterances like the second one, intuitively, it seems more likely that an error could occur in terms of the extraction of the most salient word. The reason for this assumption is the circumstance that there are firstly, two sufficiently salient words to choose from. Secondly, in presence of a less-than perfect saliency detection system where the error is partially random, more words to choose from afford the algorithm more opportunity to choose the incorrect word. The particular likelihood of such an erroneous choice depends on the structure of the 71% error as on the unknown error rate of the boundary detection. With structure of the error we mean here: Is the probability of picking a word other than the most salient one dependent on utterance length or not, and, furthermore, is this probability dependent on the number of salient words in one utterance.

Utterances per minute

Similarly to the MLU, the number of utterances per minute produced by a participant has an impact on the acquisition system: A higher number of utterances leads to more words being propagated into the robot's active lexicon.

[3]Optimally this utterance, if pronounced normally, i.e. with a small gap where the comma is located, should actually not occur as a whole in the utterance set based on which the most salient words are extracted. Yet manual inspection has shown that such utterances, and longer ones, do occur in this set.

Number of distinct words

It seems likely that the number of distinct words will have an impact on the acquisition system, if a higher numbers of distinct words leads to a higher number of distinct salient words. A higher number of salient words leads to a higher number of distinct entries in the lexicon but also to a lower number of instances per distinct word. Further research would be useful here to determine how the trade-off between a high number of instances of few distinct words versus a low number of instances of many distinct words, impacts on the performance of memory-based learners.

5.1.3 Overview of Analysis

The analysis on the utterances level is split into three parts. In subsection 5.1.4 we look at any discernible trends *within* each experiment in terms of the three measures that were just described, a so-called in-group analysis. Subsection 5.1.5 contains a cross-group analysis between the participant groups of the two negation experiments. In subsection 5.1.6, finally, a comparison on the utterance level is undertaken between the two utterance-level measures of the two negation experiments on one hand and the same measures with regards to the participants' speech within Saunders' experiments on the other.

5.1.4 In-Group Analysis

In this subsection the author focuses on the question whether any trends and statistically significant differences with regard to the measures introduced above are discernible within the speech corpus. We focus in particular on potential trends between subsequent sessions that hint towards adaptations of the speakers to the perceived cognitive abilities of the robot in terms of their linguistic behaviour.

This idea is based on the notion of recipient design (Sacks et al. 1974) which was

introduced into linguistically focused HRI research by Fischer et al. (2012). Recipient design denotes the observation that humans typically adapt their ways of speaking to the perceived needs and capabilities of the communication partner. Child-directed speech is possibly the most reported example for such adaptations, but it has also been shown to happen when the communication partner is an artificial artifact such as a robot.

Fischer et al. (2012) compare the speech of participants addressing a robot. Their study involved three different scenarios distinguished from each other by differing robotic embodiments. Their experiment was conducted based on only a single session per participant. The authors subsequently compared the effects of embodiment horizontally across different groups. Due to the single-session limitation they could not compare adaptations within the same group across several sessions ('vertical comparison'). The authors report that the majority of observed changes in their study happened in terms of measures that pertain to the interpersonal function of communication. These measure include, amongst others, the frequency of certain sentence types (imperatives, interrogative, declarative), and the frequency of use of the vocative. In order to perform latter types of measurements a more advanced analysis of our data involving the use of a parser would be needed. As these measurements are not the main focus of this thesis, we will focus on the easily obtainable subset, i.e. the aforementioned three measures, *MLU*, *u/min*, and *#dw*.

Amongst these measures Fischer et al. (2012) only observed moderate changes of the MLU across groups which were statistically significant. For multiple-session experiments such as the ones reported here it seems reasonable to expect that participants do adapt their speech with increasing familiarity with the robot. However, it is perfectly possible for such changes to go undetected if the time scale for such adaptations to take place differs significantly from the time-scale that is measured here. If participants, for example, happened to adapt their speech within the first 20 seconds of the conversation, the adap-

tation might not be visible in our measurement that spans the entire 5 minutes of each session. Two transition points of potential further adaptations are of particular interest: Firstly the transition from session 1 to session 2 which is marked by the robot starting to speak. Secondly, the transition from session 3 to 4 within the prohibition experiment, the transition of there being prohibited objects to there not being any such objects.

Rejection Experiment

utterance length (MLU) Across all sessions no obvious trend of uniform changes in MLU is discernible (see table B.15). The mean MLU based on the 50 MLUs of all sessions is 3.12 with a standard deviation of 0.72. Some participants started with a relatively low MLU in session 1, and increased it in session 2 (e.g. *P10*: 1.6 → 1.9), whereas others started with comparatively high MLUs in the first session and lowered them within the second (e.g. *P12*: 3.8 → 2.8). In order to test for statistical significance of these changes, we compared the set of utterance lengths of each session with the same set in the subsequent session. Only few changes between two subsequent sessions are statistically significant (see table B.15). Furthermore it is not the case that the largest or most significant changes within sessions happen between sessions 1 and 2. For 4 participants it is the case that their largest change in MLU happens between these two sessions, although only one of which is statistically significant though. Yet for another 4 participants the largest change in this measure happens between

Table 5.4: *Comparison of mean length of utterance between session 1 and 5 - Rejection experiment; standard deviation in brackets; * = p <.01, **= p < .05, †= p < .20*

	s1	s5	T
P01	2.22 (1.33)	2.62 (1.36)	1.57†
P04	4.69 (2.58)	4.02 (2.97)	1.47†
P05	2.78 (1.70)	2.64 (1.63)	0.71
P06	3.77 (2.29)	4.03 (2.23)	0.92
P07	3.24 (2.49)	3.10 (2.27)	0.61
P08	3.22 (2.20)	3.17 (2.07)	0.18
P09	3.76 (2.25)	3.05 (2.19)	2.59**
P10	1.63 (1.02)	1.83 (1.18)	0.61
P11	4.02 (2.75)	2.85 (2.04)	3.33*
P12	3.87 (2.34)	2.48 (1.26)	4.55*

the sessions 4 and 5.

When comparing the MLUs of each participant from the first session with the same measure for their last session, 3 participants reduced the mean utterances length in a statistically significant way and another 2 show a tendency ($p < 0.8$) to adapt their MLU. All of these changes are changes towards the 'global' MLU of 3.12 (cf. table 5.4). That is, those participants who start off with comparatively high MLUs and do adapt their utterance length (*P04*, *P09*, *P11*, *P12*), decrease their MLUs towards the 'global' mean, whereas those who start with comparatively low MLUs and who do adapt their MLUs increase their MLU towards this mean (*P01*). Interestingly the MLUs of all but one 'non-adapters' lie between those of the adapters.

Negative utterances There are no discernible trends for negative utterances only, neither when comparing subsequent sessions with each other, nor when comparing the sessions 1 and 5. See table B.15 for percentual changes and significance markers of these.

Number of distinct words (#dw) With regard to this measure we see a considerable variance between different participants, spanning from only 6 distinct words in the 1st session of participant P10 to 155 in the 3rd session of participant P09 (see table 5.2). Generally there seems to be a partial correlation between the number of produced utterances per minute and the number of distinct words, i.e. participants who talk more tend to not just repeat the same words over and over again but use different words over the course of time. But the correlation is certainly far from being perfect, as we can see by comparing the second session of participant P01 with the first session of participants P04. Though producing nearly the same number of utterances, P04 uses close to four times as many distinct words as compared to P01. In order to test for statistical significance of vertical changes in this measure across the 5 sessions, we cannot perform a t-test for each participant as

there is only one measure per participant and session (as opposed to the utterance length discussed in the last paragraph). Therefore we pool the $\#dw$ measures of all participants within each session and look for significant (vertical) trends for the group as a whole. As the variance in $\#dw$ between participants is considerable, and extreme data points distort the mean towards their respective values, extreme values have to be excluded. In the following participants are considered extreme with regard to their $\#dw$, whose $\#dw$ is either lower than $Q1-1.5*IQR$ or higher than $Q3+1.5*IQR$, with $Q1$, and $Q3$ being the first and third quartile respectively, and $IQR = Q3 - Q1$ being the interquartile range[4]. Based on this method the participants P01 and P10 were excluded due to their respective $\#dw$ values being below the $Q1 - 1.5 * IQR$ limit. No participant's $\#dw$ was above $Q3 + 1.5 * IQR$.

Table 5.5: *Development of mean #dw - Rejection experiment; T^X: without P01 and P10; entry in the T/T^X column between session x and session x+1: t-value resulting from comparison of #dw measures of session x with those of session x+1; * = p <.05*

	$mean^X$ (sd^X)	T^X	mean (sd)	T
s1	116.5 (19.85)	2.38*	95.9 (46.96)	0.94
s2	91.75 (21.75)	0.22	78.7 (33.54)	0.15
s3	88.25 (39.18)	0.05	76.2 (42.89)	0.05
s4	87.38 (37.53)	-0.25	75.2 (41.92)	0.12
s5	91.50 (28.54)		77.5 (38.8)	

Table 5.5 shows the development of the means of this measure for the remaining 8 participants across the 5 sessions. Only the transition from session 1 to session 2 is statistically significant. There we see a significant drop in the distinct number of words by on average 20% for the participants measured. This trend seems to continue in a less significant manner in the subsequent sessions until there is a marginal rise between sessions 4 and 5. For participants *P01* and *P10* which were excluded

[4]This is one method to mark the outer limits of the whiskers in a boxplot. Alternatively, one could choose the mean of $\#dw$ across one session minus one standard deviation as lower limit, and the mean plus standard deviation as upper limit, by which we would mark one participant more as outlier as compared to the IQR method. For there are only 10 participants, the decision was made in favour of the IQR method in order to lose as few data points as possible.

as outliers and both of which start with very few distinct words the opposite seems to hold true: when transitioning from session 1 to 2, they increase the number of distinct words (cf. table 5.2). Most remarkably *P10* more than triples the number of distinct words during this transition from 6 to 27, just to subsequently level in on around 25 distinct words in the sessions that follow. Due to there only being one value for *#dw* per participant and session, it is impossible to say if these changes have statistical significance for single participants that start with a rather restricted vocabulary.

Summarily we can say that all participants but one (*P09*) which start of with medium to high numbers of distinct words in session 1, reduce this number in subsequent sessions. Conceiving of these participants as a group, this reduction is statistically significant at the transition point between session 1 and 2. *P01* and *P10*, both of which start initially with very few distinct words, increase this number between session 1 and 2 and in both cases this change is the percentally largest of all changes in this measure between all sessions.

Negative utterances There are no significant trends when only considering the distinct negative words. This is the case when both considering all participants and under exclusion of participants *P01* and *P10*.

Utterances per minute (u/min) Table B.17 gives an overview of the percentual changes in *u/min*. With regard to this measure no obvious trend can be determined. While the difference in communicativeness between some participants is remarkable, just over a factor 10 between the least talkative (*P10*) and the most talkative participant (*P07*), the IQR method used in the last paragraph to identify outliers does not mark any participant as outlier. The percental changes in *u/min* are less marked than those observed with the *#dw* measure. Despite a slight drop of the mean across the sessions,

Table 5.6: *Development of mean u/min for all participants - Rejection experiment; entry in the T column at session X: t-value resulting from comparison u/min measures of session x-1 with same measures of session x*

	mean (sd)	T
s1	23.34 (10.24)	
		0.24
s2	22.33 (8.47)	
		0.24
s3	21.35 (9.79)	
		-0.02
s4	21.44 (9.49)	
		-0.03
s5	21.58 (9.82)	

there is no discernible pattern in terms of the direction of change that would hold for any subgroup of participants. Furthermore the biggest percentual changes do not generally happen between session 1 and 2, they are fairly distributed across the sessions: 3 between session 1 and 2, another 3 between session 4 and 5, and 2 between session 2 and 3 and session 3 and 4 each.

Negative utterances When considering the frequency of negative utterances only, no significant trends are discernible (cf. table B.17[5]).

Prohibition experiment

Utterance length (MLU) Similarly to the vertical comparison of MLUs within the rejection experiment no clear pattern can be observed in terms of changes of MLU across the sessions (cf. table B.18). Analoguous to the observations made for the rejection experiment it is neither the case that the largest percentual changes in this measure happen from the first to the second sessions, nor is it the case that the statistically significant changes would concentrate between these two sessions. The same observation can be made when comparing the sessions 3 and 4, those sessions between which the treatment stops. When comparing the first with the last sessions in this experiment, we see less of a change in MLU compared to one we have seen in the rejection exper-

[5]The percentages of change are numerically very high at times. This is due to the fact that the underlying base, i.e. the number of negative utterances per minute, is rather small, such that a single additional negative utterance has a large impact upon the nu/min measure.

iment: only P13 has a significantly lower MLU in session 5 as compared to session 1, and two further participants, P14 and P22, show a tendency of such a change (cf. table 5.7). P14 ends up with a higher MLU though than he or she has started with.

Table 5.7: *Comparison of mean length of utterance between session 1 and 5 - Prohibition experiment; standard deviation in brackets; * = $p < .01$, † = $p < .20$*

	s1	s5	T
P13	3.99 (2.59)	2.98 (1.76)	3.87*
P14	3.03 (1.52)	3.4 (2.16)	1.37†
P15	3.7 (2.47)	3.53 (2.35)	0.68
P16	3.63 (2.55)	3.87 (2.65)	0.91
P17	3.03 (1.76)	3.04 (2.2)	0.03
P18	3.74 (2.22)	3.6 (2.42)	0.57
P19	3.6 (2.55)	3.32 (2.19)	1.1
P20	3.68 (2.72)	3.68 (2.34)	0.02
P21	2.79 (1.77)	2.62 (1.86)	0.72
P22	3.26 (2.11)	2.9 (2.16)	1.42†

Negative utterances Also when focusing on negative utterances only we cannot discern any significant trends in MLU change (cf. table B.18).

In summary it may be said that nothing of interest can be asserted about trends of the MLU for participants of the prohibition experiment.

Number of distinct words (#dw) Generally, i.e. when considering all distinct words, no clear tendency is apparent (cf. table B.19). Half of the percentally biggest changes of this measure happen between the second and third session, yet none of the changes in *#dw* between any two sessions is statistically significant nor could be said to show a strong tendency in any direction.

Distinct negative words (#dnw) When focusing on the number of distinct negative words only, a tendency of one negation word being dropped between session 3 and 4 can be observed (cf. table 5.8). The word-level analysis in section 5.2 will show which particular word was dropped.

Table 5.8: *Distinct negative words - Prohibition experiment; sid: session id; t-values in T column between sessions x and x+1 pertain to #dnw measures of all participants between the two sessions; *= $p < .2$*

sid	mean (sd)	T	sid	mean (sd)	T	sid	mean (sd)	T	sid	mean (sd)	T
s1	4.9 (2.33)		s2	4.4 (1.43)		s3	4.7 (1.57)		s4	3.6 (2.01)	
		0.58			-0.45			1.36*			0.51
s2	4.4 (1.43)		s3	4.7 (1.57)		s4	3.6 (2.01)		s5	3.2 (1.48)	

Utterances per minute (u/min) The comparison of the utterance per minute measurement across sessions yields no significant trends (cf. table B.20). Albeit 5 out of 10 of the biggest changes in this measure happen between session 1, $u/min : 26.52$ (sd 8.03), and session 2, $u/min : 28.32$ (sd 8.62), this change is far from being statistically significant ($t = -0.48$).

Table 5.9: ***Negative utterances per minute - Prohibition experiment**: t-values in the T column between sessions x and x+1 pertain to comparison of #dnw measures of all participants between the two sessions; *= $p < 0.05$*

	mean (sd)	T
s1	4.08 (2.89)	
		-0.53
s2	4.66 (1.88)	
		0.14
s3	4.53 (2.24)	
		2.56*
s4	2.29 (1.63)	
		-0.41
s5	2.61 (1.82)	

Negative utterances per minute (nu/min) This picture changes drastically when we focus on the frequency of negative utterances only. There we can see with virtually every participant a significant drop of the *negative utterances per minute* between session 3 and 4 (cf. tables 5.11 and 5.9). This is the transition point, the treatment ends with session 3, i.e. there are no forbidden objects any more in session 4. This removes for participants the necessity to prohibit the robot. Evidently this has had a direct impact on the production of negative utterances.

Table 5.10: ***Utterance-level measures for negative utterances in rejection experiment.*** *All numbers refer to the participant with the id noted in the top row and session number in the first column. Abbreviations: sX: session nr. X, # nw/# nu: total number of negation words/negative utterances uttered by participant, # ndw: number of unique negation words, MLU: mean length of utterance , nw/min / nu/min: negation words / negative utterances per minute*

		P01	P04	P05	P06	P07	P08	P09	P10	P11	P12
s1	d (s)	272	168.4	308.6	178.3	380.9	290.4	301.9	275.5	298.3	300.5
	# nw	1	15	25	11	30	22	18	0	25	14
	# nu	1	12	24	10	28	21	18	0	21	13
	# dnw	1	5	4	3	4	3	5	0	5	3
	MLU	2	4.4	1.8	3.7	3.2	3.5	4.5	0	5.6	3.8
	nw/min	0.2	5.3	4.9	3.7	4.7	4.5	3.6	0	5	2.8
	nu/min	0.2	4.3	4.7	3.4	4.4	4.3	3.6	0	4.2	2.6
s2	d (s)	285.4	196	305.2	293.2	303.4	259.8	306.6	287.6	298.8	312.7
	# nw	3	23	21	13	35	18	21	2	41	17
	# nu	3	19	20	12	34	17	18	2	29	14
	# dnw	2	4	3	2	4	3	4	2	4	3
	MLU	2.7	4.9	3	3.5	4.1	4.5	6.2	2.5	4.2	2.8
	nw/min	0.6	7	4.1	2.7	6.9	4.2	4.1	0.4	8.2	3.3
	nu/min	0.6	5.8	3.9	2.5	6.7	3.9	3.5	0.4	5.8	2.7
s3	d (s)	307.9	297.1	249.9	318	307.8	296.8	299.2	290.8	306.6	302.5
	# nw	5	32	10	18	37	15	23	2	18	11
	# nu	5	23	10	17	35	15	20	2	15	8
	# dnw	2	4	3	4	5	4	6	2	3	1
	MLU	3.4	5.3	1.2	4.6	3	2.3	6.2	3	4.4	4
	nw/min	1	6.5	2.4	3.4	7.2	3.0	4.6	0.4	3.5	2.2
	nu/min	1	4.6	2.4	3.2	6.8	3.0	4	0.4	2.9	1.6
s4	d (s)	319.2	265.5	213.2	329.9	307.8	301.5	300.4	289.4	300.2	303.5
	# nw	8	26	12	17	25	21	14	3	21	18
	# nu	8	20	12	17	24	20	14	3	20	14
	# dnw	2	2	5	4	5	2	6	1	5	2
	MLU	1.8	4.6	3.2	3.9	3.2	3.2	5.1	4.7	3.4	2.6
	nw/min	1.5	5.9	3.4	3.1	4.9	4.2	2.8	0.6	4.2	3.6
	nu/min	1.5	4.5	3.4	3.1	4.7	4	2.8	0.6	4	2.8
s5	d (s)	305.9	220.4	269.1	307.1	314.5	324.5	301.3	290.2	319.5	299
	# nw	3	36	19	12	40	11	17	3	9	14
	# nu	3	28	16	12	40	11	15	3	6	13
	# dnw	2	5	3	3	3	2	3	1	3	2
	MLU	3.7	4.7	2.9	4.4	2.7	3.9	3.4	2	3.7	3.3
	nw/min	0.6	9.8	4.2	2.3	7.6	2	3.4	0.6	1.7	2.8
	nu/min	0.6	7.6	3.6	2.3	7.6	2	3.0	0.6	1.1	2.6

Table 5.11: ***Utterance-level measures for negative utterances in prohibition experiment.*** *Numbers refer to the participant with the id noted in the top row and session number in the first column. Abbreviations: sX: session nr. X, # nw/# nu: total number of negation words/negative utterances uttered by participant, # ndw: number of unique negation words, MLU: mean length of utterance, nw/min / nu/min: negation words / negative utterances per minute*

		P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
s1	d (s)	301.8	317.8	311.8	332.3	303.8	301.1	318.9	300.4	359.3	319.3
	# nw	22	0	61	29	16	36	22	24	33	5
	# nu	19	0	52	25	14	36	20	23	20	4
	# dnw	7	0	8	5	4	7	5	6	3	4
	MLU	4.8	0	4.5	5.6	2.7	4.4	3.8	3.9	3.6	4.8
	nw/min	4.4	0	11.7	5.2	3.2	7.2	4.1	4.8	5.5	0.9
	nu/min	3.8	0	10	4.5	2.8	7.2	3.8	4.6	3.3	0.8
s2	d (s)	324.2	305.7	301.4	307	310.7	296	308.4	308.3	317.6	312.1
	# nw	43	12	45	34	18	19	29	34	36	15
	# nu	36	8	37	32	16	18	24	30	24	15
	# dnw	5	3	6	6	3	6	4	5	4	2
	MLU	4.4	2.4	5.2	4.8	2.1	4.7	4.4	5.7	3.1	3.1
	nw/min	8	2.4	9.0	6.6	3.5	3.9	5.6	6.6	6.8	2.9
	nu/min	6.7	1.6	7.4	6.3	3.1	3.6	4.7	5.8	4.5	2.9
s3	d (s)	297.4	332.5	294.6	326.8	302.3	309.1	316.2	302.6	306.1	315.8
	# nw	7	9	49	34	26	25	29	33	25	19
	# nu	7	8	45	28	25	23	25	30	24	18
	# dnw	3	3	7	6	5	7	4	5	4	3
	MLU	4.4	3	5.0	5.0	1.9	4.9	4.7	5.3	2.4	2.7
	nw/min	1.4	1.6	10	6.2	5.2	4.9	5.5	6.5	4.9	3.6
	nu/min	1.4	1.4	9.2	5.1	5	4.5	4.7	5.9	4.7	3.4
s4	d (s)	307.6	319.9	308.5	316.1	314.7	314.8	301.7	298.6	301.1	316.2
	# nw	5	1	30	28	8	15	6	21	9	7
	# nu	5	1	24	25	8	14	6	20	8	7
	# dnw	3	1	5	6	1	6	3	6	2	3
	MLU	3.8	1	5.0	5.5	1.9	5.1	4.7	4.3	4	3.9
	nw/min	1	0.2	5.8	5.3	1.5	2.9	1.2	4.2	1.8	1.3
	nu/min	1	0.2	4.7	4.7	1.5	2.7	1.2	4.0	1.6	1.3
s5	d (s)	306.7	380.3	312.4	320.1	293.4	302.1	311.2	303.2	306.6	317.4
	# nw	6	3	32	28	7	11	10	20	16	6
	# nu	6	3	30	27	7	11	9	20	16	6
	# dnw	3	2	3	6	1	4	3	5	2	3
	MLU	5.3	1.3	2.8	5.3	1.3	4.6	4.3	4.5	2.5	2.8
	nw/min	1.2	0.5	6.1	5.2	1.4	2.2	1.9	4	3.1	1.1
	nu/min	1.2	0.5	5.8	5.1	1.4	2.2	1.7	4	3.1	1.1

Table 5.12: ***Accumulated utterance-level measures for rejection and prohibition experiment.*** *Abbreviations: sX: session nr. X, # w/# u: total number of words/utterances uttered by participant, # dw: number of distinct words, MLU: mean length of utterance based on all utterances, not on per-session MLUs, w/min / u/min: words / utterances per minute*

(a) *Rejection Experiment*

	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12
d (s)	1490.4	1147.4	1346.0	1426.5	1614.4	1473.0	1509.4	1433.5	1523.4	1518.2
# w	772	1891	1604	2952	3087	2120	2321	308	1436	976
# u	323	446	569	772	960	630	662	152	462	333
# dw	41	206	144	164	217	257	326	45	188	146
MLU	2.4	4.2	2.8	3.8	3.2	3.4	3.5	2	3.1	2.9
w/min	31.1	98.9	71.5	124.2	114.7	86.4	92.3	12.9	56.6	38.6
u/min	13	23.3	25.4	32.5	35.7	25.7	26.3	6.4	18.2	13.2

(b) *Prohibition Experiment*

	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
d (s)	1537.7	1656.2	1528.7	1602.3	1524.9	1523.1	1556.4	1513.1	1590.7	1580.8
# w	2589	1567	3397	3750	1168	3561	2916	2809	1689	2295
# u	759	469	899	999	400	965	834	701	639	770
# dw	156	121	267	404	100	365	272	190	151	215
MLU	3.4	3.3	3.8	3.8	2.9	3.7	3.5	4	2.6	3
w/min	101	56.8	133.3	140.4	46	140.3	112.4	111.4	63.7	87.1
u/min	29.6	17	35.3	37.4	15.7	38	32.2	27.8	24.1	29.2

5.1.5 Cross-Group Analysis

In this section we compare the two negation experiments, rejection and prohibition experiment, with each other. In order to compare the two experiments as a whole we will use a so-called *global measure* which is based on the superset of the underlying entities from each session, the values pertaining to our three measures, *MLU*, utterances per minute (*u/min*), and distinct number of words (*#dw*). In other words, with regards to this global

Table 5.13: ***Comparison of MLUs between* Rejection *and* Prohibition experiment**. *The two experiments are compared on a per session level as well as comparing the "global" MLU; * = p <.2*

(a) *All Utterances*

	Rejection mean (sd)	Prohibition mean(sd)	T
s1	3.31 (0.92)	3.44 (0.39)	0.41
s2	3.17 (0.74)	3.51 (0.55)	1.16
s3	3.10 (0.70)	3.39 (0.67)	0.94
s4	3.06 (0.63)	3.39 (0.38)	1.42*
s5	2.97 (0.67)	3.29 (0.41)	1.29
global	3.13 (0.65)	3.4 (0.45)	1.08

(b) *Negative Utterances Only*

	Rejection mean (sd)	Prohibition mean(sd)	T
s1	3.25 (1.61)	3.81 (1.56)	0.79
s2	3.84 (1.17)	3.99 (1.23)	0.28
s3	3.74 (1.47)	3.93 (1.28)	0.31
s4	3.57 (1.02)	3.92 (1.43)	0.63
s5	3.47 (0.8)	3.47 (1.52)	0
global	3.65 (0.91)	3.86 (1.1)	0.46

measure all entities are treated as if they were derived from a single, long session[6]. Thus, for example, the global MLUs displayed in table 5.12 differ slightly from the means of the single-session MLUs (means of means), as it is calculated not based on the latter but based on the superset of the lengths of all utterances from all sessions.

Mean-length of utterances (MLU)

Generally the MLUs of the participants within the prohibition experiment are slightly higher as compared to those of the participants within the rejection experiment, although the difference between the global measures in MLU of both experiments is not statistically significant (cf. table 5.13). The differences in means of MLU are neither statistically significant when single sessions are compared. The only exception is the difference in MLU when comparing the fourth sessions where, albeit not significant, participants within the

[6]Alternatively one could derive these global measures from the measures of the single sessions, a mean of means in the case of MLU, or, more generally speaking, derivatives of derivatives. This may still seem in order for non-derivatives such as the number of words, and may be acceptable for MLUs. Yet, with the *number of distinct words* it becomes very implausible to use the mean of the single-session measures. Using these means would largely overestimate the actual number of distinct words due to duplicate words across sessions.

prohibition experiment show a tendency to use slightly longer utterances.

For *negative utterances only* there is no significant difference between the MLUs of participants of the respective experiments.

Utterances per minute (u/min)

In terms of *utterances per minute* a pattern between the two experiments becomes apparent. The comparison of the two sets of means of this measure for each of the respective experiments (cf. table 5.14) shows that on a global level there is a tendency of participants to be more communicative within the prohibition experiment as compared to the rejection experiment.

Probably more notable is the observation that the mean of this measure starts in both experiments with approximately the same value, 23.34 vs. 22.66, the difference between the two being insignificant. Yet from session 2 onwards the means develop in opposite directions with increasing t-values such that we end up with significantly different means in session 5. While within the rejection experiment the mean drops slightly first and then levels at around 21.5, the communicativeness of the participants as measured in *u/min* within the prohibition experiment increases steadily in each session to reach approximately 30 *u/min* in session 5, nearly 40% more than the average participant's communicativeness in the rejection experiment. This is not to say that every participant in the prohibition experiment is more communicative in this last session as compared to any other participant in the rejection experiment, the in-group variance is still considerable (cf. figure 5.1). Yet, on average, participants in latter experiment talk significantly more in the last session than they did when they started. The same cannot be said about participants within the rejection experiment.

Table 5.14: *Comparison of* utterances per minute *between* Rejection *and* Prohibition experiment. *The two experiments are compared on a per session level as well as accumulatively, * = $p < .05$, † = $p < .1$, ‡ = $p < .15$*

(a) *All Utterances*

	Rejection mean (sd)	Prohibition mean (sd)	T
s1	23.34 (10.24)	22.66 (11.71)	0.14
s2	22.33 (8.47)	28.32 (8.62)	$1.58^{‡}$
s3	21.35 (9.79)	29.11 (8.57)	$1.89^{†}$
s4	21.44 (9.49)	29.23 (8.38)	$1.95^{†}$
s5	21.58 (9.81)	30.08 (7.35)	2.19^{*}
global	21.97 (9.19)	28.63 (7.8)	$1.75^{†}$

(b) *Negative Utterances Only*

	Rejection mean (sd)	Prohibition mean(sd)	T
s1	3.17 (1.73)	4.08 (2.89)	0.89
s2	3.58 (2.13)	4.66 (1.88)	1.2
s3	2.99 (1.86)	4.53 (2.24)	$1.67^{‡}$
s4	3.14 (1.3)	2.29 (1.63)	1.29
s5	3.1 (2.57)	2.61 (1.82)	0.49
global	3.2 (1.73)	3.63 (1.86)	0.53

Negative Utterances If we look at the number of utterances per minute for negative utterances only, the assertion that participants spoke more in the prohibitive experiment can generally not be made. Only in session 3 there is a tendency ($p < 0.15$) of participants within the prohibition experiment to speak more compared to those of the rejection experiment. In the last two sessions, i.e. after the removal of the prohibition task, the opposite seems to be the case. On average, albeit with a much weaker tendency, participants in the rejection experiment uttered more negative utterances.

Number of distinct words (#dw)

Participants in the prohibition experiment used tendentially more distinct words as compared to the participants of the rejection experiment. Although, when the distinct words of all 5 sessions are accumulated ('global level'), this difference is not statistically significant (cf. table 5.15). If all participants are taken into account, there is a strong tendency for this difference in session 2 and a weaker tendency in session 3. Yet if we exclude the most extreme participants of both experiments, *P01* and *P10*, as we did for

Table 5.15: ***Comparison of*** **number of distinct words** ***between rejection and prohibition experiment****. The two experiments are compared on a per session level as well as accumulatively. Furthermore t-numbers are given for the case when participants P01 and P10 are excluded from the analysis (T^X); † = $p < .05$, * = $p < .1$, ** = $p < .15$, ‡ = $p < .2$*

(a) *All Utterances*

	Rejection mean (sd)	Prohibition mean(sd)	T	T^X
s1	95.9 (46.96)	121.3 (46.60)	1.21	0.29
s2	78.7 (33.54)	110.6 (46.36)	1.76*	1.14
s3	76.2 (42.89)	112.6 (51.34)	1.72**	1.11
s4	75.2 (41.92)	108.1 (55.34)	1.5‡	0.9
s5	77.5 (38.8)	105.2 (51.79)	1.35‡	0.71
global	173.4 (87.67)	224.1 (101.99)	1.19	0.46

(b) *Negative Utterances Only*

	Rejection mean (sd)	Prohibition mean(sd)	T	T^X
s1	3.3 (1.7)	4.9 (2.33)	1.75*	1.12
s2	3.1 (0.88)	4.4 (1.43)	2.45†	1.96*
s3	3.4 (1.51)	4.7 (1.57)	1.89*	1.31
s4	3.4 (1.78)	3.6 (2.01)	0.24	0.31
s5	2.7 (1.06)	3.2 (1.48)	0.87	0.33
global	6.2 (3.12)	7.2 (2.39)	0.8	0.16

the vertical comparison in section 5.1.4, the statistical tendencies disappear. This means that the extremely low number of distinct utterances of participants *P01* and *P10* have a non-negligible impact on the mean.

Number of distinct negative words Participants from the prohibition experiment tend to produce one more negation word as compared to participants of the rejection experiment during the first three sessions. This difference is statistically significant for session 2 and shows a strong tendency in sessions 1 and 3. If we exclude participants *P01* and *P10* from the data set the statistical significance disappears, only session 2 shows a strong tendency. Furthermore, and independent of the two outliers, this difference disappears in sessions 4 and 5. Upon exclusion of said participants the mean of the rejection experiment in session 4 rises to 3.88 (sd: 1.64) and to 3 (sd: 0.93) in session 5. For the global measure no difference in number of distinct negative words can be shown.

Table 5.16: ***Accumulated utterance-level measures for negative utterances and words only.*** *Abbreviations: sX: session nr. X, # w/# u: total number of words/utterances uttered by participant, # dnw: number of distinct neg. words, MLU: mean length of utterance, w/min / u/min: words / utterances per minute*

(a) *Rejection Experiment*

	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12
d (s)	1490.4	1147.4	1346.0	1426.5	1614.4	1473.0	1509.4	1433.5	1523.4	1518.2
# nw	20	132	87	71	167	87	93	10	114	74
# nu	20	102	82	68	161	84	85	10	91	62
# dnw	4	6	5	5	9	5	12	2	10	4
MLU	2.6	4.8	2.5	4.1	3.2	3.5	5.2	3.1	4.3	3.2
nw/min	0.8	6.9	3.9	3	6.2	3.5	3.7	0.4	4.5	2.9
nu/min	0.8	5.3	3.7	2.9	6.0	3.4	3.4	0.4	3.6	2.5

(b) *Prohibition Experiment*

	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
d (s)	1537.7	1656.2	1528.7	1602.3	1524.9	1523.1	1556.4	1513.1	1590.7	1580.8
# nw	83	25	217	153	75	106	96	132	119	52
# nu	73	20	188	137	70	102	84	123	92	50
# dnw	7	4	10	11	6	9	6	9	5	5
MLU	4.5	2.4	4.5	5.2	2.1	4.7	4.3	4.8	3	3.1
nw/min	3.2	0.9	8.5	5.7	3	4.2	3.7	5.2	4.5	2
nu/min	2.8	0.7	7.4	5.1	2.8	4	3.2	4.9	3.5	1.9

5.1.6 Comparison with Saunders et al. (2012)

In this subsection a comparison is undertaken between participants' speech recorded during the two negation experiments and the speech of Saunders et al.'s (2012)[7] participants. The architecture used within Saunders' experiments is very similar to the architecture employed for the experiments reported about in this thesis. The major difference between the two architectures is the addition of the motivation system described in section 3.2 and the newly designed behavioural system (cf. section 3.3). As outlined there, the major design rationale

[7]For reasons of simplicity we will refer to the authors in the following just as "Saunders" without any intention to deny recognition to any of the other authors.

of this system is to be as convincing as possible in terms of expressing these motivational states or, in other words, to maximise the congruence between the internal motivational state and the outwardly expressed behaviour. Is is worth emphasizing that there is no difference between the employed degrees of freedom by the robot or its physical appearance compared to Saunders et al. (2012). We therefore do believe that the two systems as well as the experimental setup are similar enough to make a linguistic comparison possible without running the danger of these measures being confounded by factors other than the presence of the motivation system, the independent variable so to speak. Yet there are three more differences between the two experiments that have to be mentioned for the sake of a fair comparison:

1. The duration of a session within the negation experiments is approximately twice as long as the duration of one of Saunders' sessions.

2. The objects in Saunders' experiment were coloured and had varying sizes.

3. The instructions to the participants in the two negation experiments draw attention to the fact that the robot has preferences towards the objects and that it will express these.

In essence the generated behaviour of Saunders' system is very similar to the hypothetical behaviour of our behavioural system if we would remove the motivational system. This, taken together with the fact that the experimental tasks of our rejection experiment and Saunders' experiment were identical, causes us to believe that speech from Saunders et al.'s (2012) experiment is suitable for a comparison. Our motivation behind this comparison is the evaluation of the impact of the motivational system, accompanied by congruent behaviour, upon the speech of naïve participants. In the following we will use ANOVA analyses as the speech of three groups will be compared with each other: participants from

Table 5.17: ***Accumulated utterance-level measures for participants' speech from Saunders' experiments (Saunders et al. 2012).*** *The listed participant ids are compatible to the ones used in latter publication. Abbreviations: sX: session nr. X, # (n)w/# (n)u: total number of (negative) words/utterances uttered by participant, # d(n)w: number of distinct (negative) words, MLU: mean length of utterance, nw/min / nu/min: (negative) words / utterances per minute*

(a) *All Utterances*

	M02	F05	M03	F01	F02	M01	F03	F06	F04
d (s)	726	671.5	646.1	568.6	603.6	663.1	501.4	706.8	621.3
# w	610	1044	540	627	1124	811	847	1420	1035
# u	204	272	188	159	244	237	262	386	323
# dw	40	121	76	52	96	70	106	168	112
MLU	3	3.8	2.9	3.9	4.6	3.4	3.2	3.7	3.2
w/min	50.4	93.3	50.1	66.2	111.7	73.4	101.4	120.5	100
u/min	16.9	24.3	17.5	16.8	24.3	21.4	31.4	32.8	31.2

(b) *Negative Utterances Only*

	M02	F05	M03	F01	F02	M01	F03	F06	F04
d (s)	726	671.5	646.1	568.6	603.6	663.1	501.4	706.8	621.3
# nw	0	1	9	8	0	0	1	60	6
# nu	0	1	8	7	0	0	1	54	5
# dnw	0	1	3	2	0	0	1	8	3
MLU	0	8	5.2	5.3	0	0	5	5.4	6
nw/min	0	0.1	0.8	0.8	0	0	0.1	5.1	0.6
nu/min	0	0.1	0.7	0.7	0	0	0.1	4.6	0.5

the rejection experiment, participants from the prohibition experiment, and participants from Saunders' experiment. Table 5.17 gives an overview of our measurement in Saunders' experiment, complete tables can be found in section B.2 of the appendix (tables B.3 and B.4).

Mean length of utterances (MLU)

The comparison of the mean length of utterances between the three experiments shows no significant difference (cf. table 5.18).

Table 5.18: *Comparison of* MLU *between negation experiments and Saunders' experiment;* $\star$*: due to only 2 data points no mean was calculated;* $* = p < 0.001$, $\dagger = p < 0.85$, $\ddagger = p < 0.01$

(a) *All Utterances*

	Saunders mean (sd)	Rejection mean (sd)	Prohibition mean(sd)	F
s1	3.64 (0.61)	3.31 (0.92)	3.44 (0.39)	0.52
s2	3.62 (0.55)	3.17 (0.74)	3.51 (0.55)	1.38
s3	3.53 (0.54)	3.11 (0.71)	3.39 (0.67)	1.07
s4	3.49 (0.66)	3.06 (0.63)	3.39 (0.38)	1.53
s5	3.43 (0.75)	2.97 (0.67)	3.29 (0.41)	1.41
global	3.52 (0.54)	3.13 (0.65)	3.40 (0.45)	1.27

(b) *Negative Utterances Only*

	Saunders mean (sd)	Rejection mean (sd)	Prohibition mean(sd)	F
s1	11.8 (6.49)	3.61 (1.2)	4.23 (0.84)	14.32^{*}
s2	5.1 (1.41)	3.84 (1.17)	3.99 (1.23)	1.58
s3	4.34 (2.32)	3.74 (1.47)	3.93 (1.28)	0.24
s4	$n/a^{\star}$	3.57 (1.02)	3.92 (1.43)	0.4
s5	4.88 (1.18)	3.47 (0.8)	3.47 (1.53)	$2.23^{\dagger}$
global	5.82 (1.12)	3.65 (0.91)	3.86 (1.1)	$9.22^{\ddagger}$

Negative Utterances Only When only considering negative utterances the ANOVA indicates a significant difference in MLU in the first session and in the global measure. But this difference has to be considered with caution. The underlying basis of negative utterances within Saunders' experiment is extremely small. The average of negative utterances per session are 153 (sd 8.6), 187.8 (sd 57.1), and 15.2 (sd 4.82) for the rejective, prohibitive, and Saunders' experiment respectively. The on average 15 negative utterances are a very small basis for the calculation of the mean, such that one very long utterance has the potential to skew the mean considerably. Indeed the very large MLU of 19 for M03 in Saun-

ders' first session (cf. table B.4) is based on one single utterance[8]. As the automatic utterance segmentation has never been formally tested and could potentially produce undeservedly long utterances from time to time, the utterance length of any single utterance has to be considered very critically. Therefore we regard the significance of this difference in MLU skeptically, especially as it is not repeated in any other session.

Table 5.19: ***Comparison of* #dw *between negation experiments and Saunders' experiment.*** *g.: global, * = $p < 0.001$, † = $p < 0.01$, ‡ = $p < 0.05$*

(a) *All Utterances*

	Saunders mean (sd)	Rejection mean (sd)	Prohib. mean(sd)	F
s1	61.88 (25.8)	95.9 (47)	121.3 (46.6)	4.45‡
s2	51.56 (20)	78.7 (33.5)	110.6 (46.4)	6.61†
s3	49.33 (19.5)	76.2 (42.9)	112.6 (51.3)	5.77†
s4	43.56 (21.8)	75.2 (41.9)	108.1 (55.3)	5.45‡
s5	43.55 (18.8)	77.5 (38.8)	105.2 (51.8)	5.78†
g.	93.44 (39.1)	173.4 (87.7)	224.1 (102)	6.07†

(b) *Negative Utterances Only*

	Saunders mean (sd)	Rejection mean (sd)	Prohib. mean(sd)	F
s1	1 (1.77)	3.3 (1.7)	4.9 (2.33)	8.73†
s2	1.11 (1.96)	3.1 (0.88)	4.4 (1.43)	11.96*
s3	1.11 (1.62)	3.4 (1.51)	4.7 (1.57)	12.73*
s4	0.78 (1.99)	3.4 (1.78)	3.6 (2.01)	6.23†
s5	0.78 (1.3)	2.7 (1.06)	3.1 (1.52)	8.39†
g.	2 (2.58)	6.2 (3.12)	7.2 (2.39)	9.66*

Number of distinct words (#dw)

With regards to the measure *number of distinct words* the statistics indicate a significant difference between the three experiments. This is mainly due to the lesser number of distinct words in Saunders' experiment (cf. the much lower significance levels in table 5.15 where only rejection and prohibition experiments are compared). This significant difference can be possibly explained to a large degree with the much shorter duration of

[8]It should be noted that an MLU of 19 is not impossible though. We doubtfully investigated a similarly long utterance of one participant in the prohibition experiment, suspecting a fault of the automatic boundary detection. It appeared that this participant spoke extremely fast in the given situation without any audible pause. Thus the automatically established boundaries seemed adequate in that case.

Saunders' sessions. As participants do not constantly repeat themselves, i.e. they use at least in part different words as they continue to talk, one would expect a higher number of distinct words in experiments where participants talk for longer. The relationship between duration and the number of distinct words is naturally not linear: some words, especially function words such as articles, prepositions, or personal pronouns will be used repetitively, but also object words typically recur when talking about the same object for a second or third time. Subsequently one would not expect twice as many distinct words to occur in a session that lasts twice as long, but one would nevertheless expect an increase in this measure that is solely caused by the longer duration. Subsequently the approximately 85% increase of this measure between Saunders' experiment and the rejection experiment is not unexpected. It does not appear reasonable to make any further definite statements with regards to the impact of the independent variable, i.e the presence of a motivational system, on this measure.

Number of distinct negation words The yet more significant increase in the number of distinct negation words between Saunders' experiment and the rejection experiment, a more than 300% increase, does indicate an influence of the independent variable. It seems impossible though to tease apart the influence of the two factors, presence of a motivation system and longer duration of the session, upon the statistical significance of the difference in this measure. A look on a measure, that takes account of duration, such as *utterances per minute* appears more suited to shed light on this issue.

Utterances per minute (u/min)

When comparing the *u/min* measure of all utterances of the three experiments we can

Table 5.20: *Comparison of* u/min *between negation experiments and Saunders' experiment;* † = $p < 0.01$, * = $p < 0.05$, ** = $p < 0.1$, ‡ = $p < 0.2$

(a) *All Utterances*

	Saunders mean (sd)	Rejection mean (sd)	Prohibition mean(sd)	F
s1	24.86 (8.01)	23.34 (10.2)	26.52 (8.03)	0.32
s2	25.31 (5.04)	22.33 (8.47)	28.32 (8.62)	1.54
s3	24.4 (6.68)	21.35 (9.79)	29.11 (8.57)	2.11‡
s4	21.9 (6.99)	21.44 (9.49)	29.23 (8.38)	2.67**
s5	23.12 (7.31)	21.58 (9.82)	30.08 (7.35)	2.97**
global	24.06 (6.48)	21.97 (9.19)	28.63 (7.8)	1.83‡

(b) *Negative Utterances Only*

	Saunders mean (sd)	Rejection mean (sd)	Prohibition mean(sd)	F
s1	0.6 (1.19)	3.17 (1.73)	4.08 (2.89)	6.31†
s2	0.96 (1.86)	3.58 (2.13)	4.66 (1.88)	8.83†
s3	1.1 (1.86)	2.99 (1.86)	4.53 (2.24)	6.98†
s4	0.73 (1.85)	3.14 (1.3)	2.29 (1.63)	5.45*
s5	0.48 (0.9)	3.1 (2.57)	2.61 (1.82)	4.93*
global	0.74 (1.48)	3.2 (1.73)	3.63 (1.86)	7.77†

see an increasing tendency of the participants in the prohibition experiment to speak more than in the rejection experiment, with Saunders' participants being in between (cf. significance levels in table 5.14). The tendency reaches near-significance in session 5 (F=2.97, p = 0.069).

Negative utterances only When considering the *u/min* measure for negative utterances only (cf. table 5.20), the difference between the two negation experiments and Saunders' experiment is yet more striking. Participants in the rejection experiment produce on average more than four times as many negative utterances as compared to participants in Saunders' experiment (3.2 vs. 0.74). For the prohibition experiment the ratio reaches nearly factor 5 (3.63 vs. 0.74) on average, with it being even bigger in the first three sessions. Here is a very strong indication that the motivation system impacts the linguistic behaviour of participants: it provokes the production of

negative utterances.

When considering the prohibition experiment in isolation, one might argue that the fact that participants were asked to physically prohibit the robot caused them to engage in linguistic means of prohibition. So this argument could attempt to explain the sudden surge of negative utterances by emphasising the prohibitive nature of their actions, which were triggered by the experimenter telling them to act in this way. This argument collapses, when the rejection experiment is taken into account: there was no mention of prohibition, neither in the written nor in the oral instructions on part of the experimenter. In terms of instructions, the only difference to the ones given in Saunders' experiment is that participants were told that the robot has preferences towards the objects, that the robot would communicate these preferences to them in some way, and, furthermore, that they should present all objects to the robot in order to teach it their names, not minding too much about its potential rejection of some of them.

Within Saunders' experiment on average only 3.08 % of all utterances produced are negative utterances, whereas in the rejection and prohibition experiment this fraction rises to 14.57% and 12.68% respectively. This also indicates that the difference in the number of distinct negation words observed in the previous section were caused by many more negative utterances being produced. Assuming that participants do vary the words they use when speaking about similar topics or when performing similar linguistic deeds, an increase in the distinct number of negative words appears a natural outcome of a strong increase in the number of negative utterances being produced.

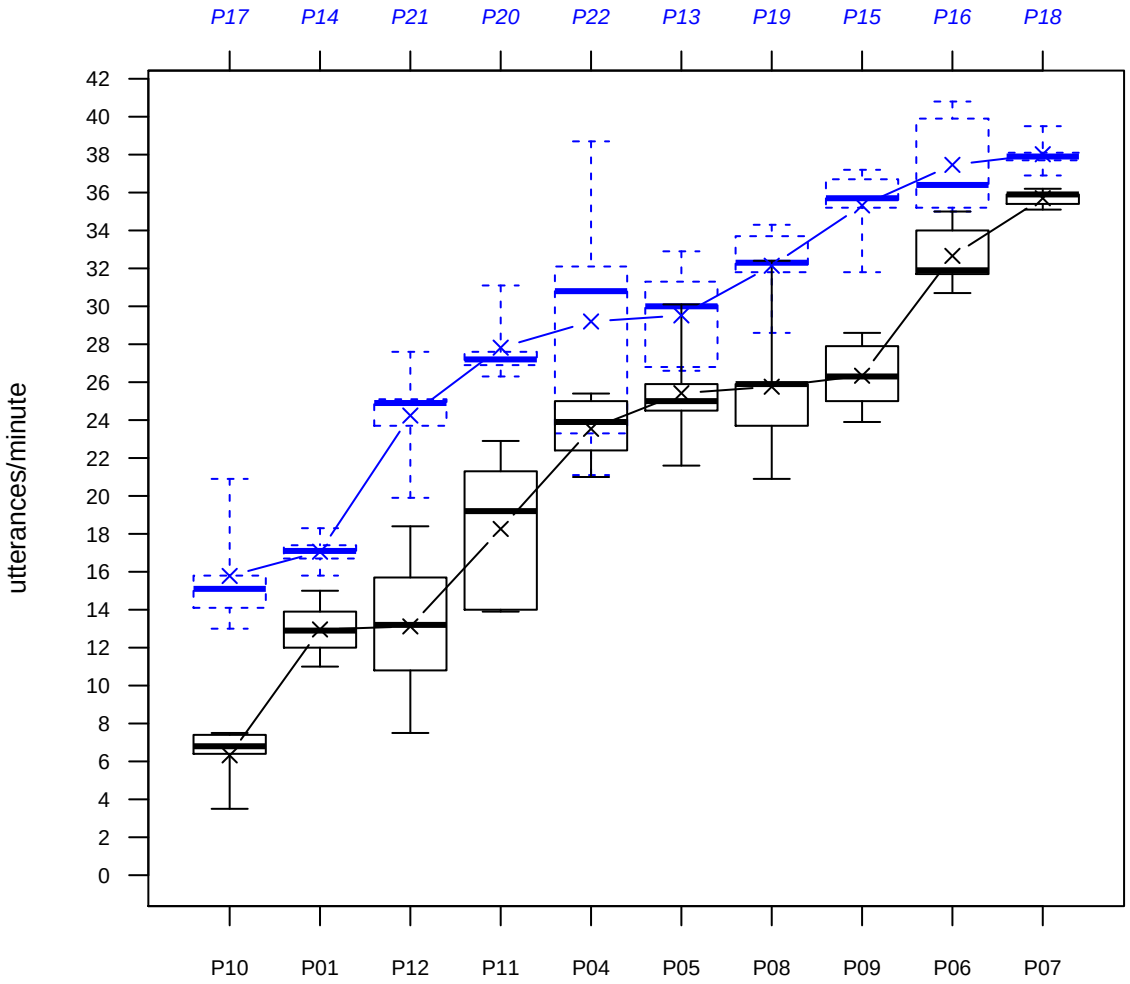


Figure 5.1: ***Communicativeness of participants within each experiment****. Displayed are the means of* utterances per minute *by the symbol '×' and connected through lines for each participant and experiment as well as boxpots of the same coefficient (see also table 5.12). Participants are ordered within each experiment in ascending order with regards to said coefficient, i.e. their average degree of communicativeness across the sessions. Upper (blue) line with '×'s: means of* utterances per minute *per participant within prohibition experiment. Lower (black) line with '×'s: mean of* utterances per minute *per participants within rejection experiment. The boxplots visualize the median, maximum, minimum, 1st and 3rd quartiles of* utterances per minute *for each participant. The upper line of blue, dashed boxplots corresponds to participants within prohibition experiment. The lower line of black, solid boxplots corresponds to participants within the rejection experiment. Participant ids for the prohibition and rejection experiments are displayed on the top and bottom horizontal axes respectively.*

5.2 Human: Word Level

There are at least two reasons why an analysis at the word level is of interest to the research reported in this thesis. The first, probably less evident reason has to do with measures of so called *interpersonal function* which deliver clues with regard to the impact of a robot's behaviour upon the involvement of its human interlocutor. The second reason for looking at participants' production frequencies of single words, is that salient words form the basis of our acquisition algorithm and, as a whole and separately for each participant, constitute the productive lexicon of the robot.

Interpersonal Function

In the previous section the analytical focus was on utterance level characteristics of the participants' speech such as the mean length of utterance (MLU), the utterances per minute, and the number of distinct words. In the arsenal of linguistic measures put forward by Fischer et al. (2012) the latter two measures pertain to measures of *linguistic verbosity*, whereas the first measure, MLU, pertains to the measures of *utterance complexity* (cf. section 5.0.2). Fischer et al.[9] list a third class: measures of *interpersonal function*. The latter ought to tell us something about "the degree with which speakers involve their communication partner" (Fischer et al. 2012). This is of interest to us because the motivation behind our attempt to create an 'emotionally convincing' behaviour system is the hypothesis that participants would be prompted to engage in intent interpretations similar, or possibly even identical, to those that have been observed within CDS. Furthermore, in a second move, negative intent interpretations are hypothesized to potentially serve as a basis for the acquisition of certain negation types. We suspect that the production of intent in-

[9]We will at times for ease of reference refer to the authors just as *Fischer*, especially when using the possessive ending *'s*. No disrespect towards the other authors is intended.

terpretations by a participant implies a considerable degree of him or her being involved with the robot. If this is so, we would expect to be able to detect this involvement using measures from Fischer's *interpersonal function* category.

Many of the measures listed in Fischer et al. (2012) are not easy to obtain as they involve the detection of sentence mood or the use of the infinitive. In order to detect and count these one needs either a manual analysis of the entire speech corpus or the use of an advanced parser that is able to detect said grammatical categories. Luckily the authors list three other measures, two of which are at the word level and the forth, by pure chance, coincides with one particular negation type, which will be introduced in the next section. The three measures are:

- the frequency of the personal pronouns *you*, *I*, and *we*
- the frequency of the vocative, i.e. the robot's name *Deechee*
- the use of understanding checks, in English typically tag questions: *don't you*, *isn't it*, etc

High counts of these measures indicate, according to Fischer et al. (2012), a rather personal relationship between speaker and hearer. Impersonal relationships, according to this reasoning, would be reflected in a less frequent use of personal pronouns, vocatives, and tag questions, and would furthermore result in comparatively high frequencies of impersonal constructions as reflected by the use of the German *man* (*one*).

As the impersonal *one* is used far less in conversational English as compared to conversational German it is doubtful if this measure is of much use in experiments with native English speakers. We will for this reason not consider *one*.

Furthermore we would like to argue that the usage of *you/Du* in English and German are sufficiently different such that numbers reflecting the frequency of their use cannot be

directly compared across the two languages. The reason for this is the following: The use of the impersonal *man (one)* is a perfectly valid choice in colloquial German that does not carry the heavy connotations of its British equivalent such as alleged or aspired membership to the upper class or an 'academic mode of conversation'. In contrast, the use of *one* is not or to a much lesser degree used in colloquial British English, and has the aforementioned connotations. These connotations render it much less of a choice compared to a German speaker's choice of *man*. Conversely, this means that the choice to use *Du* (*you*) is a proper choice in German, the speaker could have chosen *man* (*one*) instead without the fear of sounding presumptuous. Yet 'choosing' *you* is not much of a choice for an English speaker, due to the lack of real alternatives, it rather amounts to the default. For these reasons we think that in English *you* is not as indicative of a speaker's involvement with a hearer as Fischer claims this to be the case in German. We will include *you* in our analysis nevertheless by way of stating its frequency and coverage, but will not make implications with regards to its indicativeness with regard to speaker involvement. We hope that attempts will be made in the future to work out measures for interpersonal function that apply to English in a way that Fischer's measures seem to apply to German.

Tag questions as further indicators of involvement, cannot be identified on a word level as the involved negation words such as *isn't* in *isn't it*, or *don't* in *don't you* are not unique to tag questions. Tag questions will therefore be covered in section 5.3, where they, by pure coincidence, happen to be classed as a separate negation type.

Word learning - Negation Words

As this research focuses on the acquisition of negation we will in the following pay particular attention to negation words in the frequency list. The selective frequency tables in this chapter also contain the 'usual suspects' that one would expect to be frequent in a noun-

learning scenario and as they were reported in Saunders et al. (2012): object words such as *star*, *moon*, *heart*, or *square* that relate directly to the symbols attached to the boxes whose names participants were supposed to teach the robot. We will consider negation words in the context of these object words, especially when comparing the word frequencies to those in Saunders' experiment. Apart from using accumulated frequency tables of all words, frequency tables of prosodically salient words will be employed. Only prosodically salient words form the actual basis of the word acquisition algorithm and are in this regard more important than general, "prosodically blind" word frequencies.

Overview of section

Subsection 5.2.1 presents a comparative analysis of the two negation experiments on a word level based on accumulative word frequency tables which contain the words of all participants and sessions. We will also use the British National Corpus on spoken English as a measuring stick to put our measurements in the context of word frequencies as they are found in ordinary conversations. Our use of accumulative frequency tables does not adjust for the communicativeness of participants. As a result the frequency tables will be skewed towards the word frequencies of the more talkative participants. In order to alleviate this potential point of criticism we will give each participant equal influence on the accumulated word rankings in adjusted ranking tables in subsection 5.2.2. There, a voting algorithm is employed to determine the word ranks. This method ensures that the ranked word lists of participants, no matter how talkative or taciturn, 'get an equal vote' in determining the ranks of the accumulated list. This algorithm ignores the word frequencies alltogether and only considers the rank of each word within a single participant's word frequency list. By doing so, we lose the frequency information, but ensure that each participant has the same impact on the global result. In subsection 5.2.3 we will finally compare

the accumulated word ranks and frequencies with those calculated from the transcripts of Saunders' experiment (Saunders et al. 2012). This comparison seeks to determine if the presence of the motivational system produces a measurable impact in terms of word frequencies and ranks.

Notation The following expressions will be used in the analyses:

- *Top 10* and *top 20* refer to the 10 and 20 most frequent words in the frequency tables respectively.
- *Coverage* denotes the share of words (tokens) that a particular word form (type) accounts for in the corpus. Example: The word form *no* has one entry in frequency table 5.22 and its percentage is given as 2.39. In this case we will say that *no* has a coverage of 2.39% or covers 2.39% of the entire corpus.

5.2.1 Accumulated Word Frequencies

In order to have a reference point with wich to compare the frequency lists derived from our corpus we consider it helpful to look at general word frequencies in spoken British English. Table 5.21 lists the top ranking 20 words of spoken English from the BNC (Leech et al. 2001), which was retrieved from Companion Website for: Word Frequencies in Written and Spoken English by Geoffrey Leech et. al (2001). We will refer to this corpus as *BNCS* from here on.

Notice that the frequency table on the basis of which table 5.21 was generated does not have exactly the same structure as our frequency tables due to the fact that the tables from Leech et al. (2001) use a part-of-speech tagging, such that the same word occurs several times in the table if it has more than one grammatical function[10]. We checked the

[10]A quick search of entries starting with *a*, for example, yielded 6 entries: one for *a* in isolation when it acts as determiner (freq.: 18637), another entry for *A / a* in isolation with *a* acting as letter (freq.: 405), and four more entries where *a* is part of adverbial constructions such as *a bit* (freq.: 496). All frequencies

10 top-ranking words for lower-rank entries containing these words and their impact on their ranks. Some of the entries such as *that* which has two major grammatical functions, determiner (freq.[11]: 14252) and conjunction (freq.: 7246), would move up a few ranks, if it was merged into one entry. Generally our analysis did not indicate that merging of all entries for one particular word would result in a fundamentally different frequency table with regard to the top ranking words. Evidently the 20 most frequent words in spoken British English are dominated by function or closed class words: personal pronouns *I*, *you*, *it*, *we*, prepositions *of*, *to*, *that*, determiners *a*, *the*, and the negation word *n't* (*not*). Furthermore the auxiliary verbs *do*, *have*, *is*, and *'s* (*is*), but also the so called interjection *yeah* can be found in these top ranks. *No*, also classified as interjection in the BNC, ranks on place 41.

Table 5.21: *20+1 most frequent words of BNCS. Abs. frequencies based on 1 million words.*

rank	word	freq	%
1	the	39605	3.96
2	I	29448	2.94
3	you	25957	2.60
4	and	25210	2.52
5	it	24508	2.45
6	a	18637	1.86
7	's	17677	1.77
8	to	14912	1.49
9	of	14550	1.46
10	that	14252	1.43
11	n't	12212	1.22
12	in	11609	1.16
13	we	10448	1.04
14	is	10164	1.02
15	do	9594	0.96
16	they	9333	0.93
17	er	8542	0.85
18	was	8097	0.81
19	yeah	7890	0.79
20	have	7488	0.75
...	...	...	...
41	no	4388	0.44

Rejection Experiment

Table 5.22 lists the frequencies of the 10 most frequent and other selected words from the accumulated speech of all participants in the rejection experiment (cf. table B.5 in the appendix for the complete listing). We will refer to this set of words in the following as *Rejection Corpus (RC)*. When comparing table 5.22 to the BNCS frequency table (5.21) we find the following differences:

per million.

[11]all total frequencies are counts per million

Table 5.22: ***Word-frequencies of all words in rejection experiment****. Listed are the ten most frequent words within said experiment across all participants and sessions. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot. See table B.5 in the appendix for the complete listing.*

rank	word	cnt	%	*rank*	word	cnt	%	rank	word	cnt	%
(1)	you	1245	7.13	*(35)*	Deechee	127	0.73	*(93)*	didn't	11	0.06
(2)	the	983	5.63	*(38)*	not	118	0.68	*(94)*	didn't (2)	10	0.06
(3)	like	579	3.31	*(41)*	box	110	0.63	*(95*	pyramid	9	0.05
(4)	a	475	2.72	*(42)*	Deechee (2)	103	0.59	*(93)*	isn't	11	0.06
(5)	this	471	2.7	*(44)*	triangles	99	0.57	*(97)*	moons	7	0.04
(6)	no	417	2.39	*(54)*	don't (2)	69	0.4	*(100)*	rectangle	4	0.02
(7)	one	396	2.27	*(58)*	crescent	58	0.33	*(101)*	won't	3	0.02
(8)	square	337	1.93	*(65)*	sad	43	0.25	*(100)*	smiling	4	0.02
(8)	do	337	1.93	*(68)*	shape	39	0.22	*(101)*	can't	3	0.02
(9)	to	311	1.93	*(69)*	happy	38	0.22	*(101)*	pyramids	3	0.02
(11)	moon	283	1.62	*(69)*	nice	38	0.22	*(103)*	wouldn't	1	0.01
(12)	heart	279	1.6	*(70)*	favourite	37	0.21	*(102)*	doesn't	2	0.01
(14)	triangle	254	1.45	*(71)*	target	36	0.21	*(102)*	doesn't (2)	2	0.01
(15)	circle	231	1.32	*(75)*	hearts	30	0.17	*(103)*	couldn't	1	0.01
(17)	don't	200	1.15	*(78)*	arteen	27	0.15	*(103)*	wasn't	1	0.01
(21)	circles	190	1.09	*(86)*	know	18	0.1	*(103)*	weren't	1	0.01
(23)	squares	180	1.03	*(91)*	smile	13	0.07	*(103)*	can't (2)	1	0.01
(24)	yes	179	1.02	*(92)*	rectangles	12	0.07	*(100)*	haven't	4	0.02

- *you* is the top-ranking word in the RC (coverage 7.13%) as opposed to *the* in the BNCS. In the BNCS *you* covers 2.6%.

- *like* is part of the top 10 in the RC and covers 3.31% of all words in this corpus. In the BNC it is distributed across 5 ranks [12]. The combined coverage of all these entries in the BNCS is 0.370%.

[12]format: [*rank (PoS) freq.*]: 96 (Prep) 1762, 138 (Verb) 1070, 163 (Adv) 784, 1200 (Conj) 61, and 2946 (Adj) 19

- *no* is part of the top 10 and covers 2.39% in the RC. In the BNCS it is distributed across 5 entries[13]. The combined coverage in the BNCS is 0.558%. In the RC it ranks higher than any object-related word, i.e. object labels and terms that denote properties of objects such as colour, size, etc.
- *square*, an object label, is part of the top 10 within the RC
- the ranks 10 to 20 in the RC contain many object labels
- *don't* is amongst the top 20 in the RC. It has a second phonetic variant on rank 54 and the combined coverage of both variants is 1.55%. It has no entry in the BNCS as it seems to split there into *do* and *n't*. As *n't* is a postfix for other auxiliary verbs as well, we cannot calculate rank or coverage of *don't* in the BNCS.
- If *Deechee*, the vocative, would have been transcribed into a single phonetic form, it would rank amongst the top 20 in the RC (combined count: 230)

In summary we can say that the dominance of function words in the top 20 of frequency tables for spoken English as indicated in the BNCS is weakened in the frequency table of the speech collected during our word-learning experiment. There we find object labels amongst the top 20, which might not be very surprising within a word-learning experiment. Yet *you* occurs more frequently than on average in spoken English, possibly indicating an increased involvement of our participants with the robot as compared to 'the average conversation'. Another surprisingly well above-average frequent word to be found in the RC is *like*, stemming from utterances such as "do you like the heart". In a similar vein the vocative represented by *Deechee* is highly frequent and is another indicator for heightened involvement. Furthermore *no* is more than 4 times more frequent in the RC than in the

[13] 2 entries with *no* in isolation and 3 where it is part of an expression. Format: [*rank (PoS) freq.*]: 41 (Int) 4388, 136 (Det) 1102, 1495 (Pron) 45, 2141 (Adv) 29, 3044 (Adv) 18

BNCS indicating that the hypothesised elicitation mechanism did indeed work. We will see more detail about the source of *no* in the pragmatic analysis in section 5.3.

Nevertheless we still find non-personal function words such as the determiners *the* and *a* as well as the preposition *to* amongst the most frequent words. The picture changes dramatically when we look at the frequency of words, that are prosodically salient.

Prosodically Salient Words Before going into detail it is worth reflecting what a change in rank between the two tables, the general frequency table 5.22 and the frequency table of salient words 5.23 means. The ranks in the tables reflect the comparative percentage of the respective words proportional to the total number of words. If a word drops in the ranks, it means that the number of times where it was prosodically emphasized by the speakers is below average compared to other prosodically salient words. In terms of notation we will use *salient* to mean *prosodically salient* in the remainder of this section. Besides we will use the abbreviation *RCS* for the corpus of salient words derived from the speech of all participants of the rejection experiment.

Table 5.23 lists the frequencies and ranks of the ten most frequent salient words as well as the ranks and frequencies of the words which were already listed in the general frequency table (table 5.22).

The differences to the RC frequency table is striking: function words other than *no* and *it* disappeared from the top 10. The within RC highly frequent determiners *a* and *the* moved down to the ranks 47 and 41 respectively with only 2.32% and 1.73% of their occurrences being salient. The top ranks, which were filled with functions words in the RC, are replaced by object labels in the RCS. *No*, which was on rank 6 in the general frequency table and had a coverage of 2.39% in the RC words moved up to rank 2 in the RCS and makes up for 4.64% of all salient words, only to be topped by *square* whose coverage is 0.33

Table 5.23: ***Word-frequencies of salient words in rejection experiment****. Listed are the ten most frequent salient words within said experiment across all participants and sessions. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot. See table B.6 in the appendix for the complete listing.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	square	259	4.97	*(24)*	don't	52	1	*(46)*	do	12	0.23
(2)	no	242	4.64	*(27)*	box	45	0.86	*(47)*	rectangles	11	0.21
(3)	triangle	206	3.95	*(29)*	crescent	39	0.75	*(47)*	a	11	0.21
(4)	heart	198	3.8	*(31)*	are	36	0.69	*(50)*	pyramid	8	0.15
(5)	moon	184	3.53	*(32)*	sad	35	0.67	*(53)*	isn't	5	0.1
(5)	circle	184	3.53	*(33)*	target	31	0.59	*(54)*	rectangle	4	0.08
(6)	like	167	3.2	*(34)*	favourite	27	0.52	*(55)*	pyramids	3	0.06
(7)	circles	140	2.69	*(35)*	arteen	25	0.48	*(55)*	moons	3	0.06
(8)	squares	126	2.42	*(36)*	not	24	0.46	*(55)*	smiling	3	0.06
(9)	it	123	2.36	*(37)*	hearts	22	0.42	*(56)*	can't	2	0.04
(10)	yes	119	2.28	*(38)*	to	20	0.38	*(56)*	won't	2	0.04
(11)	one	111	2.13	*(38)*	happy	20	0.38	*(57)*	didn't (2)	1	0.02
(13)	this	95	1.82	*(39)*	shape	19	0.36	*(57)*	couldn't	1	0.02
(18)	Deechee	76	1.46	*(39)*	nice	19	0.36	*(57)*	doesn't	1	0.02
(19)	Deechee (2)	68	1.3	*(41)*	the	17	0.33	*(57)*	didn't	1	0.02
(21)	triangles	62	1.19	*(46)*	smile	12	0.23	*(57)*	haven't	1	0.02
(23)	you	53	1.02	*(46)*	don't (2)	12	0.23				

higher than that of *no*. Moreover 58.03% of all occurrences of *no*'s are salient. Also notice that *like* is still a member of the ten most frequent words despite its rank dropping by 4. Furthermore the vocative *Deechee* moved up in the ranks. The sum of the frequencies of both phonetic variants is 130, thus if *Deechee* would have been transcribed using the same phonetic variant, the vocative would rank on 8th place, contributing to 2.76% of all salient words with 62.6% of its occurrences being salient.

Table 5.24 lists the salience rates of highly frequent words and types of words in the general frequency table. The salience rate for each word or word group is calculated by simply dividing the frequency of the (set of) words in the RCS by the frequency of this (set

of) words in the RC. As can be seen, the reason for the disappearance of articles from the top ranks is that they hardly ever are salient. The contrary is the case for object labels, which are the most salient group of words. Also noteworthy is the high salience rate of emotion words as well as the high prosodic salience of the vocative.

Table 5.24: *Prosodic salience rates of selected word groups and words - Rejection Corpus. Obj. labels:* square, triangle, heart, moon, circle, crescent, target, arteen, rectangle, pyramid, *plus plurals; emotion words:* sad, happy, smile*; demonstratives:* this, that*; pers. pronouns:* you, we, I*; articles:* a, the*; vocative:* Deechee, Deechee (2)

word (group)	saliency rate (%)
object labels	73.81
emotion words	71.28
yes	66.48
vocative	62.6
no	58.03
it	39.55
like	28.89
one	28.03
don't	23.79
demonstratives	20.31
to	6.43
pers. pronouns [14]	4.15
do	3.56
articles	1.92

Negation Words In terms of negation words it is noteworthy that the 'major players' for the rejection experiment turn out to be *no*, *don't*, and to a lesser degree *not*. All the other negation words listed at the beginning of this chapter (table 5.1), mainly auxiliary verbs with the postfix *n't* such as *isn't*, were produced extremely rarely (coverage $\ll 0.1\%$). *No* ranks high in the RCS mainly because of its general high frequency but also due to half of its productions being salient, whereas *don't* drops slightly in the ranks due to its lower, but still considerable salience rate of above 23%. The frequent and salient occurrence of *don't* within the rejection experiment took us a bit by surprise, as we expected it to occur mainly as part of prohibitive utterances. As it transpired, many participants used *don't* when asking motivation-dependent questions such as "You

[14] If expressions containing these personal pronouns such as *I've*, *we'll* etc., which are single phonetic words, are included in the count, the salience rate changes to 4.26%. The expressions found in the corpus are: *I've*, *I'll I'm*, *we've*, *we'll*, *we'd*, *you've*, *you're*, *you'd*, and *you'll*.

don't like that?". In these questions participants evidently emphasize *don't* very frequently, thereby indicating one of two possibly related things: First, emphasis on *don't* can indicate real surprise on the part of the participant and the question is a way of requesting affirmation that something is indeed not the case, with the *something* here being Deechee's liking. Second, emphasis on *don't* may indicate doubt about the speakers current assumption that Deechee indeed dislikes something. This is a somewhat weaker uncertainty than proper surprise, but in both cases, doubt or surprise, the speaker pauses after posing the question and awaits some kind of feedback[15] from the recipient, in order to determine his or her next move in the interaction game. Of course this does not necessarily mean that the speaker subsequently acts in accordance with the preferences of the conversation partner, but it seems to be important in conversations to make clear, in terms of preferences, where each of the conversation partners stands.

Interpersonal Function As can be seen in the general frequency table 5.22 *you* turns out to be the most frequent word uttered by participants, albeit not very salient. It covers 7.13% in the RC as opposed to 2.96% in the BNCS. The percentage of *you* and variants in the RC relative to the number of utterances in the experiment (5309) is 24.12%, Fischer et al. (2012) report 3%, 6%, and 2% for their three different HRI scenarios respectively[16]. But remember, that their target language was German, such that a direct cross-linguistic comparison is most probably untenable, as striking as this difference may appear to be. The two phonetic variants of *Deechee*, the vocative, have a relatively high combined salience rate (62.6%), and a combined coverage of 1.32% in the RC. Due to the

[15]Remember that feedback in conversation is not necessarily a positive move, but can also be given via the absence of any action. Cf. the CA example in section 2.1.3.

[16]The scenario closest to the rejection scenario is scenario nr. 3 where the complete iCub is used, so the 3rd of the given numbers would be the reference for comparison, if they same target language would have been used.

high salience rate its coverage in the RCS rises to 2.76% which would correspond to the 7th rank if combined. The percentage of the vocative relative to the number of utterances is 4.33%, Fischer et al. report $1\%, 4\%,$ and 16% respectively for their three scenarios. *We* covers 1.18% within RC as compared to 1.04% in the BNCS, it's percentage relative to the number of utterances is 3.92%[17] compared to $4\%, 12\%,$ and 4% respectively in Fischer's scenarios. *I* covers 0.86% in the RC compared to 2.96% in the BNCS[18]. Furthermore its percentage relative to the number of utterances is 3.33%[19] compared to $15\%, 17\%,$ and 12% in Fischer's scenario.

On the Word Distribution Figure 5.2 depicts graphically the word distributions which very roughly resemble a Zipf distribution. Zipf's Law states that the relationship between the rank of a word and its frequency corresponds to $r^\rho f = c$, with r being the rank, f being the frequency, c being a constant, and ρ being a parameter (Wyllys 1981)[20]. ρ is typically within the interval $[0.9, 2]$. Zipf's law has been observed to hold for various word distributions of many languages, but also other distributions related to social phenomena such as city sizes (Wyllys 1981). So far the word distribution of our corpus is unsurprising and comparable to other corpuses. This we consider a good thing in the sense that it indicates that our participants did not engage in an extremely untypical way of speaking.

[17] combined percentage of *we*, *we've*, *we'll* and *we'd*
[18] 2 PoS variants: pronoun (29448 occ.) and letter (193 occ)
[19] combined percentage of *I*, *I'll*, *I've*, *I'm*
[20] Determining the value of c requires slightly heavier mathematical machinery (cf. Weisstein 2013)

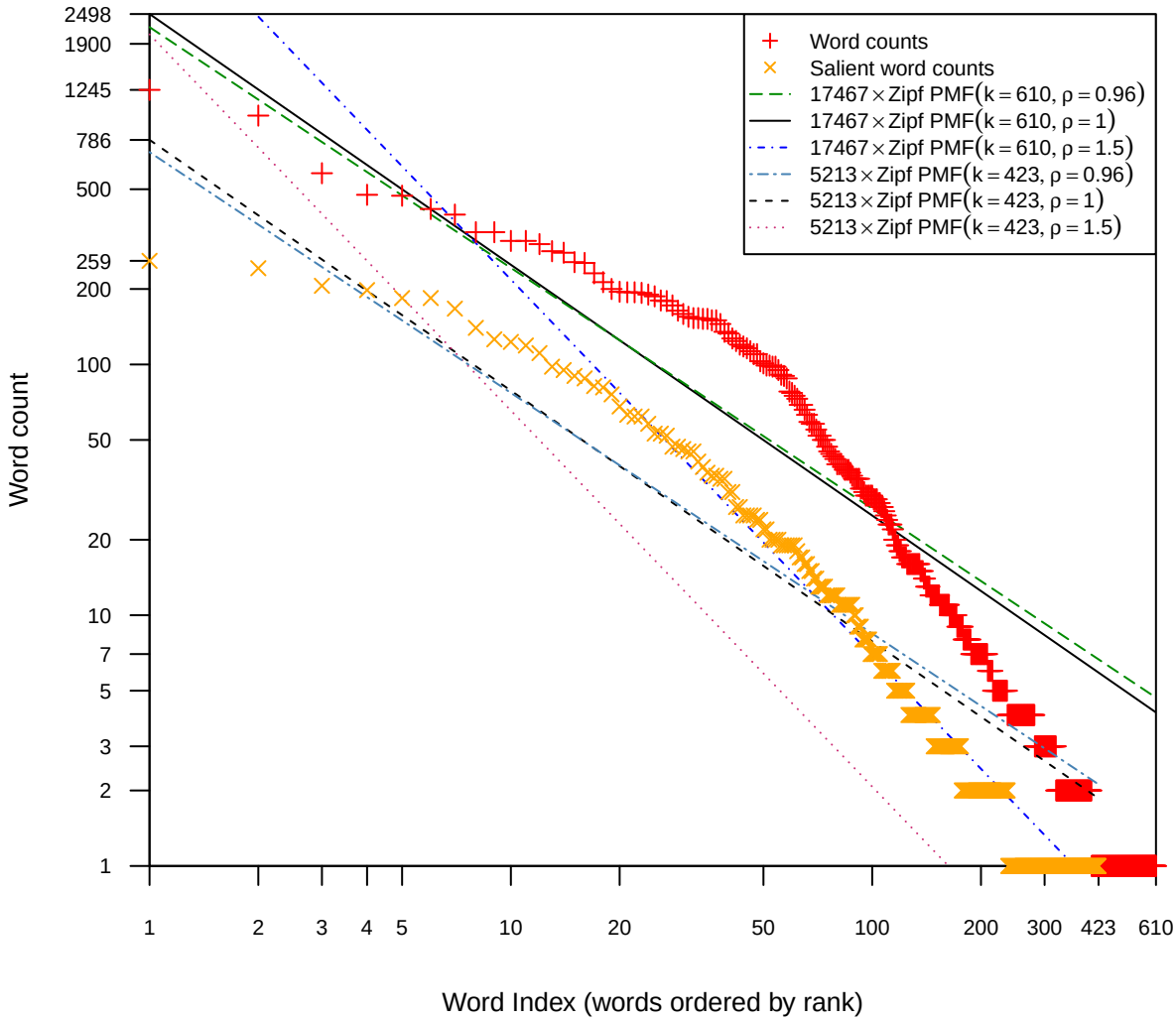


Figure 5.2: ***Word counts within the rejection experiment.*** *Displayed are the words counts for all sessions and participants of the rejection experiment on a log-log scale. The word index is plotted on the x-axis, words are indexed by rank in ascending order, i.e. falling ranks from left to right. For all words the total number of 17464 words is distributed over 613 distinct phonetic words which were produced by the participants within this experiment. For salient words only, 5213 words are distributed over 423 distinct phonetic words. Both distributions of word frequencies resemble roughly a Zipf distribution. Three Zipf probability mass functions (PMF) per distribution, multiplied with the total number of words are plotted along the word counts for comparison. Within the word distribution of all words 50% of these words are accounted for by the 24 highest ranking distinct words. Within the word distribution of salient words only 50% of these words are accounted for by the 19 highest ranking distinct salient words. For comparison, this is also the case with a Zipf PMF with* $\rho = 0.96$. *See tables 5.22 and 5.23 for top ranking and other selected words.*

It might help to imagine how a fairly untypical way of speaking, which would not result in a Zipf-like distribution, would look like. Imagine the majority of our participants would only ever have uttered the object label when showing an object to the robot and nothing else. In this case the word frequencies in such a corpus of less than 20 words, exclusively object labels, would be distributed in a manner similar to a uniform distribution. Deviations from a uniform distribution would in this case have only been caused by unequal numbers of object presentations, leading to an unequal number of word labels being uttered, leading to a slight slope in a regular, non-log-log-plot.

The fact that the fit to the Zipfian' is far from being perfect might have to do with the fact that our underlying word-basis, i.e. the number of distinct words, is comparatively small in our corpus. When considering all words, the corpus derived from the speech in the rejection experiment consists of only 613 words. When considering salient words alone, this number diminishes to 423. These comparatively small numbers of distinct words might explain why our word distributions show a considerable deviation from a Zipfian distribution for the middle ranks, the "hump" towards the right in the middle section. Furthermore the fact that the words are drawn from speech that is constrained to a very particular situational context, our word-teaching experiment, might further lead to a deviation from the Zipfian as compared to larger corpora that are derived from speech from various situational contexts. The circumstance that the very highest and very lowest ranks don't show a good fit to the Zipfian' is not unusual for word corpuses (Wyllys 1981).

Despite all imperfection of fit, our distribution is still rather Zipfian than, say, normal or uniform, in the sense that a few highest-ranking word forms[21] (types) contribute the majority of words (tokens) in the corpus and that the rank-frequency relationship is roughly inversely proportional. This leads to a exponentially rather than linearly falling slope of

[21]Our lexicon/corpus is made up of phonetic words/word forms. So in the context of our corpus, *word* and *word form* should be read as abbreviation for *phonetic word* and *phonetic word form*.

the rank-frequency plot. In the case of the "all words" corpus the highest-ranking 24 word forms, only 3.92% of all word forms, cover over 50% of all words in the corpus (red '+'-plot fig. 5.2). In the case of salient words only, already the first 19 word forms, 4.49% of all salient word forms, yield a 50% coverage (yellow '×'-plot, same figure).

Prohibition Experiment

Table 5.25: ***Word-frequencies of all words in prohibition experiment****. Listed are the ten most frequent words within said experiment across all participants and sessions. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot. See table B.7 in the appendix for the complete listing.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	you	1591	6.18	*(27)*	yes	227	0.88	*(114)*	sad	8	0.03
(2)	the	1416	5.5	*(31)*	not	198	0.77	*(116)*	smiling	6	0.02
(3)	a	962	3.74	*(32)*	crescent	184	0.72	*(116)*	haven't	6	0.02
(4)	this	956	3.71	*(33)*	circles	168	0.65	*(117)*	mustn't	5	0.02
(5)	one	722	2.8	*(43)*	Deechee (2)	132	0.51	*(118)*	won't	4	0.02
(6)	is	632	2.45	*(47)*	box	123	0.48	*(119)*	target	3	0.01
(7)	like	527	2.05	*(56)*	Deechee	103	0.4	*(119)*	hasn't	3	0.01
(8)	to	471	1.83	*(58)*	can't (2)	98	0.38	*(120)*	crescents	2	0.01
(9)	no	461	1.79	*(63)*	squares	86	0.33	*(120)*	cannot	2	0.01
(10)	it's	428	1.66	*(66)*	hearts	79	0.31	*(120)*	pyramids	2	0.01
(11)	heart	411	1.6	*(69)*	triangles	72	0.28	*(121)*	shouldn't	1	0
(12)	square	389	1.51	*(75)*	nice	59	0.23	*(121)*	moons	1	0
(13)	triangle	377	1.47	*(81)*	know	51	0.20	*(121)*	nono	1	0
(15)	do	360	1.40	*(88)*	favourite	34	0.13	*(121)*	wouldn't	1	0
(16)	moon	356	1.38	*(100)*	happy	22	0.09	*(121)*	neither	1	0
(17)	circle	332	1.29	*(101)*	smile	21	0.08	*(121)*	pyramid	1	0
(19)	shape	310	1.2	*(102)*	isn't	20	0.08	*(121)*	weren't	1	0
(26)	play	229	0.89	*(103)*	didn't	19	0.07				
(26)	don't	229	0.89	*(113)*	doesn't	9	0.04				

Table 5.25 lists the general word frequencies of the corpus stemming from the speech

produced by participants within the prohibition experiment, called prohibition corpus (*PC*) henceforth.

The ranking of word forms in this table is very similar to the word ranking of the Rejection Corpus (table 5.22), yet there are small differences. We would like to highlight the following word forms:

- *you* ranks first and covers 6.18%
- *like* is amongst the top 10 and covers 2.05%
- *no* is amongst the top 10 and covers 1.79%
- The ranks 10 to 20 are dominated by object labels
- *don't* is not amongst the top 20 any more, it's coverage is 0.89%
- Another negative, *can't*, appeared above the 0.1% mark, it's coverage is 0.38%

Prosodically Salient Words Table 5.26 gives an overview of ranks and frequencies of selected prosodically salient words originating from the corpus of salient words extracted from the speech linked to the prohibition experiment. We will henceforth refer to this corpus as *PCS*.

We can observe similarly drastic changes in terms of word ranks and frequencies between *PCS* and *PC* as between *RCS* and *RC*. Again, all but two function words disappeared from the top 10. *A* moved down to rank 53, with only 2.18% of its tokens being salient. *The* moved down to rank 35 with only 3.18% of its tokens being salient. *No* moved up to rank 4, with 59% of its tokens being salient and is one of two function words that remain in the top 10. The other one is *this*, salience rate: 24.06%. Interestingly, *ok*, that in a

Table 5.26: ***Word-frequencies of salient words in prohibition experiment.*** *Listed are the ten most frequent salient words within said experiment across all participants and sessions. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot. See table B.8 in the appendix for the complete listing.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	square	327	4.4	*(25)*	Deechee	67	0.9	*(65)*	know	9	0.12
(2)	triangle	302	4.06	*(27)*	squares	62	0.83	*(69)*	smiling	5	0.07
(3)	circle	285	3.83	*(27)*	is	62	0.83	*(71)*	target	3	0.04
(4)	no	272	3.66	*(32)*	triangles	50	0.67	*(71)*	sad	3	0.04
(5)	one	250	3.36	*(35)*	the	45	0.61	*(72)*	mustn't	2	0.03
(6)	heart	248	3.34	*(35)*	don't	45	0.61	*(72)*	crescents	2	0.03
(7)	this	230	3.09	*(37)*	hearts	43	0.58	*(72)*	cannot	2	0.03
(8)	moon	208	2.8	*(40)*	can't (2)	39	0.53	*(72)*	doesn't	2	0.03
(9)	ok	171	2.3	*(45)*	not	31	0.42	*(72)*	haven't	2	0.03
(10)	shape	159	2.14	*(47)*	do	29	0.39	*(73)*	nono	1	0.01
(11)	yes	155	2.09	*(49)*	favourite	27	0.36	*(73)*	won't	1	0.01
(12)	crescent	149	2	*(53)*	a	21	0.28	*(73)*	hasn't	1	0.01
(13)	like	134	1.8	*(54)*	to	20	0.27	*(73)*	pyramids	1	0.01
(14)	circles	129	1.74	*(54)*	nice	20	0.27	*(73)*	neither	1	0.01
(20)	Deechee (2)	83	1.12	*(57)*	it's	17	0.23	*(73)*	pyramid	1	0.01
(21)	you	79	1.06	*(59)*	smile	15	0.2	*(73)*	weren't	1	0.01
(23)	play	75	1.01	*(64)*	happy	10	0.13				
(24)	box	72	0.97	*(65)*	isn't	9	0.12				

part of speech categorization might be classed as interjection[22], and therefore a function word, moved up into the top 10. The salience rate of *ok* is an astonishing 77.73%. The remaining words in the top 10 are all object labels and *one*, which most probably stems from expressions such as "What about this one". Table 5.27 shows the salience rates for various words and word groups based on the prohibition corpora *PC* and *PCS*. As is the case in the rejection experiment object labels lead the table with a close to identical salience

[22]The grammarians don't seem to agree very much on the categorization of these 'interjections': In the BNCS *yes* is classed as interjection, whereas *okay* is classed as adverb. OED online, "the definitive record of the English language", classes *ok* as adjective, interjection, and noun (in this order), but *yes* is classed there as adverb and other things, interjection not being one of them (OED online 2013).

rates in both corpora, 73.81% and 73.49% in the RCS and PCS respectively. The group of emotion words are less salient in the PCS, 54.90% as compared to the RCS, 71.28%.

Table 5.27: *Prosodic salience rates of selected word groups and words. Prohibition Corpus - object labels:* square, triangle, heart, moon, circle, crescent, target, pyramid, *and their plurals; emotion words:* sad, happy, smile*; demonstratives:* this, that*; pers. pronouns:* you, we, I*; articles:* a, a(2), the*; vocative:* Deechee, Deechee (2)

word (group)	saliency rate (%)
object labels	73.49
yes	68.28
vocative	63.83
no	59
emotion words	54.90
can't	39.80
one	34.63
it	27.60
like	25.43
demonstratives	22.39
don't	19.65
do	8.06
to	4.25
pers. pronouns[23]	5.25
articles	2.78

Yes (RCS: 66.48%, PCS: 68.28%), *no* (RCS: 58.03%, PCS: 59%), the vocative (RCS: 62.6, PCS: 63.83), and *like* (RCS: 28.89%, PCS: 25.43%) have nearly identical salience rates in both experiments. Equally the salience rates of demonstratives (RCS: 20.31%, PCS: 22.39%), articles (RCS: 1.92%, PCS: 2.78%), and personal pronouns (RCS: 4.15%, PCS: 5.25%) are almost on a par. For reasons that we cannot explain the auxiliary verb *do* has within the PCS a salience rate of more than twice that of the RCS.

Negation Words: Prohibition vs. Rejection Experiment Similarly to what was observed in the rejection experiment *no* is high-ranking in the PCS due to a combination of high overall frequency and high prosodic salience - more than every second *no* is produced with prosodic emphasis. The salience rate of *don't* in the PC(S) dropped slightly compared to the one in RC(S) as did its general frequency resulting in an overall lower rank in the PCS.

[23]If expressions containing these personal pronouns such as *I've*, *we'll* etc., which are single phonetic words, are included in the count, the salience rate changes to 5.15%. The expressions found in the corpus are: *I've*, *I'll I'm*, *I'd*, *we've*, *we'll*, *you've*, *you're*, and *you'd*.

This observation surprised us, as we expected participants to use rather more than less *don't*s when producing prohibitive utterances such as "Don't touch the square". What we observe instead, is that another negation word, *can't*, emerged in the middle ranks. It was produced 98 times by participants and roughly 40% of these productions were prosodically emphasized to such degree that it was the most salient word of the utterance. It was typically part of prohibitive utterances such as "You can't have that one", "You can't play", "You can't have the circle", etc. Within the rejection experiment it played virtually no role. *Can't* was only produced there 4 times by any of the 10 participants and never in a prohibitive context.

Not has a marginally higher frequency in PC compared to RC, but a lower salience rate (RC(S): 20.34%, PC(S): 15.66%). For all other negation words the same observation can be made as for their role within the rejection experiment: they hardly occur. *Isn't* is the only one that comes close the 0.1% margin, occurring 20 times and with 9 of its productions having been salient.

Interpersonal Function: Prohibition vs. Rejection Experiment All percentages given in the following are relative to the number of utterances in the prohibition experiment unless stated otherwise. *You* is still the most frequent word in PC and the combined percentage of *you* and variants[24] is 21.60, a drop by −2.53% compared to RC. The percentage of the vocative *Deechee* is 3.16, a drop by −1.17% compared to RC. For *we* plus variants the combined percentage amounts to 4.49, a rise of 0.57%. The percentage of *I* and variants within the PC is 3.13, which is a drop by −0.2% compared to RC.

If we exclude *you* as indicator of interpersonal function for the reasons given in subsection 5.2, and sum up the percentual differences of each indicator of interpersonal function

[24]With "variants of *you*" we mean grammatical compound expressions containing *you*, but which amount to a single phonetic word. In the PC these are *you've*, *you'd*, and *you're*.

between PC and RC, the result is a drop by -0.8 in the use of personal pronouns and vocatives in the prohibition experiment compared to the rejection experiment. This is a drop of -6.91% relative to the sum of values of these indicators from the RC, 11.58% of utterances within the RC include one of these personal pronouns or a vocative.

Without statistical analysis it is impossible to say if this is a significant value, but it does not seem high enough to us to draw the definite conclusion that participants in within the prohibition experiment were significantly less involved with the robot compared to the rejection experiment.

On the Word Distributions Figure 5.3 depicts graphically the word distributions of the PC and PCS, which roughly resemble a Zipf distribution. The distributions are very similar to the distributions of the RC and RCS, hence we see no need for a separate discussion and refer to section 5.2.1 where the distributions of RC and RCS are discussed.

5.2.2 Adjusted Accumulated Word Frequencies

The merger of corpora of single participants into one big corpus, which forms the basis of the analysis in the previous section, has one major disadvantage: as every word has the same weight in determining percentual changes of words and word groups, the resulting tendencies are not necessarily tendencies of 'the average participant'. Participant P07, the most communicative participant of the rejection experiment for example contributes more than 10 times as many words to the merged corpus as participant P10, the least communicative participant. Thus P07 has 10 times more influence on our 'global' measure than P01, because all of these measures are based on word counts. In other words 'the average participant' is in terms of any word-based measure not really an average participant, but reflects to a much larger degree the linguistic characteristics of P07 than those of P10.

Determining group-ranks by voting In order to ensure that the ranking in the rank-frequency tables is indeed representative of all participants we employ the *Ranked Pairs (RP)* algorithm (Tideman 1987), a voting algorithm which, based on a set of ballots containing entries sorted by rank, calculates one ranked list. The comparison of this list with our accumulative frequency tables should give us an indication if the rankings given there represents the ranking of all participants. If the ranked words list produced by *RP* is considerably different from the rankings in our frequency tables, we know that the influence of the more communicative participants biased the group ranking considerably.

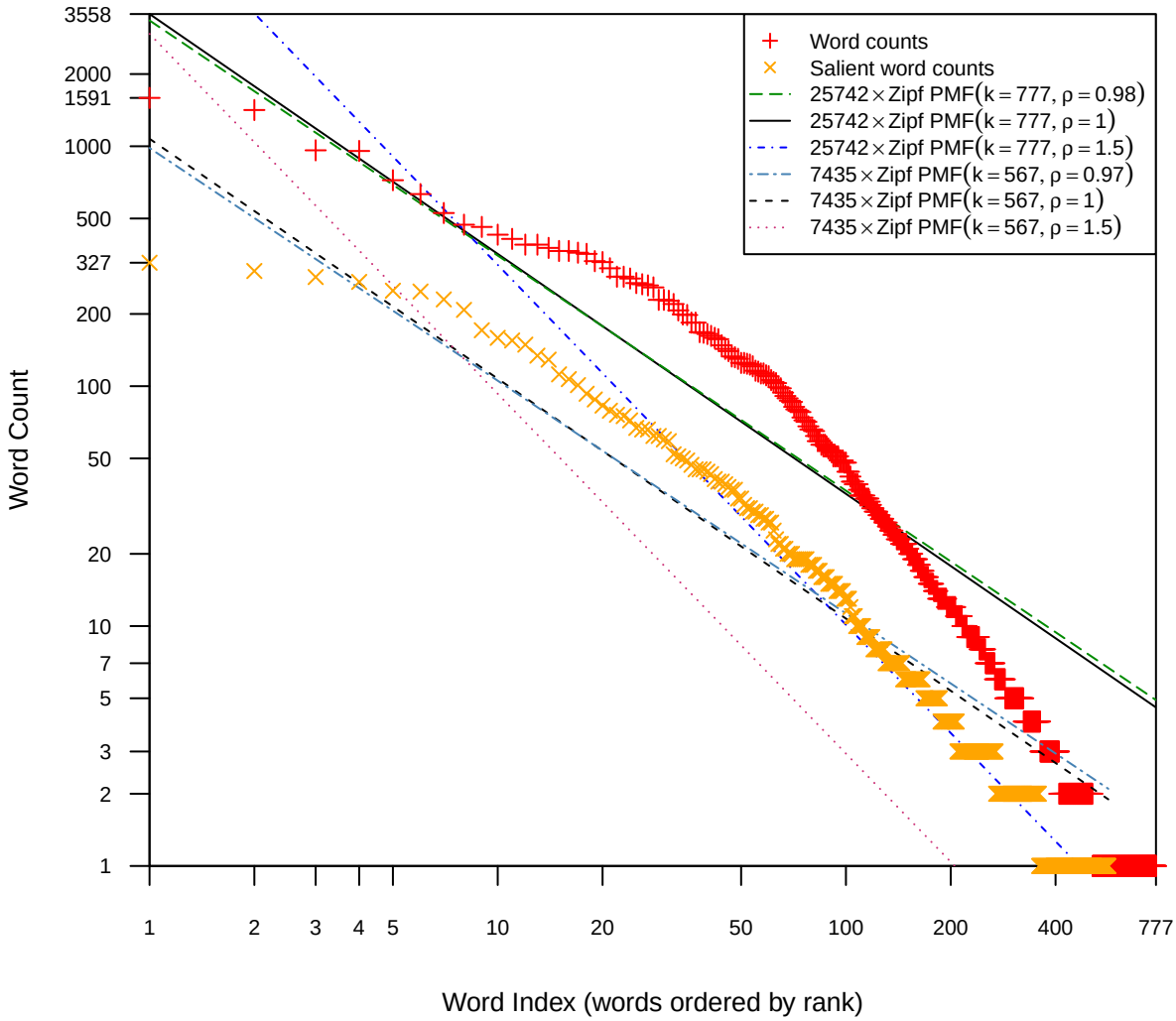


Figure 5.3: ***Word counts within the prohibition experiment.*** *Displayed are the words counts for all sessions and participants of the prohibition experiment on a log-log scale. The word index of the words is plotted on the x-axis - words are indexed by rank in ascending order, i.e. falling rank from left to right. For all words the total number of 25745 words is distributed over 778 distinct phonetic words which were produced by the participants within this experiment. For salient words only 7434 words are distributed over 567 distinct phonetic words. Both distributions of word frequencies resemble roughly a Zipf distribution. Three Zipf probability mass functions (PMF) for each distribution, multiplied with the total number of words are plotted along the word counts for comparison. Within the word distribution of all words 50% of these words are accounted for by the 24 highest ranking distinct words. Within the word distribution of salient words only, 50% of these words are accounted for by the 22 highest ranking distinct salient words. For comparison, this is also the case with a Zipf PMF with 778 entities for* $\rho = 0.98$ *and with a Zipf PMF with 567 entities for* $\rho = 0.97$*. See tables 5.25 and 5.26 for top ranking and other selected words.*

If the rankings produced by *RP* are to a large degree identical to those on the accumulated rank-frequency lists we can be fairly sure that a statistical analysis based on the rank-frequency lists of single participants would not give us a considerably different ranking of words. The disadvantage of a ranking algorithm as compared to a statistical comparison is the circumstance that ranking algorithms do not give us any quantification of the differences in the measures within a cross-group analysis. Therefore the percentual differences given in the last section will have to be considered with a certain scepticism despite any possible affirmation of correctness in terms of rankings via the *Ranked Pair* voting.

The main idea behind the application of a voting algorithm to a frequency list of words derived from a corpus is as follows: Conceive of each participant as a voter who produces a ranked list of words ordered by their frequency. By interpreting this list as a ballot as used within a voting procedure the problem of determining an accumulative list of word ranks based on a set of subordinate lists of such word ranks can be transformed into a problem of determining a list of winners in an election based on subordinated lists where preferred winners are ordered by preference, most preferred candidate first. The chosen algorithm assigns to every voter the same weight in determining the overall outcome of the vote and virtually ignores all frequency information. This eliminates eventual bias through largely differing frequencies between voters. The outcome of this voting procedure is a ranked list of all available entries on the ballots, which in our case corresponds to a ranked word list that is representative of the individual ranked word lists as a whole.

For reasons of computational complexity we did not run the *RP* algorithm on the complete frequency lists turned ballots. Instead the decision was made to include the n highest-ranking entries on each ballot, such that the combined coverage of these n entries reaches at least 50% coverage for each participant's corpus. As n was determined by the participant's

frequency list which required the most entries to reach the 50% coverage, the coverage for most participants is considerably larger. For the RC, RCS, PC, and PCS n is 23, 23, 30, and 31 respectively.

Outcome of the voting process The resulting ranked lists when applying the *ranked pairs* voting procedure on the ballots from each of the four corpora are given in table 5.28. In the following four paragraphs the ranking in the rank-frequency tables of the four corpora are compared to the word rankings as calculated by the *Ranked Pair* algorithm.

Rejection Corpus (RC): Ranks by vote vs. Ranks by frequency When comparing the voting-based ranking in table 5.22 (see also complete table B.5 in appendix) with the frequency-based ranking (table 5.28) no major differences can be established. The words in the top 10 are basically identical, only *to* moved down to rank 16 in the vote-based table and *moon* moved up to rank 10. Furthermore there are some minor rank-swaps within the top 10: *no* moved up by 1, *square* moved up by 2, and *like* moved down by 1. When extending the scope to the top 20, we can see that the object labels are still there, with *heart* having moved down by 3, and *don't* having moved down by 5.

Table 5.28: ***Adjusted accumulated word rankings in both negation experiments****. Listed are the 25 top-ranking words of all words and of salient words only within each experiment. The ranking within each list results from conceiving of the frequency-ordered word lists for each participant as voting ballots that are ordered descendingly with regard to word-frequency. The frequency itself is not considered though. This approach eliminates the greater influence of very talkative participants on the accumulated rankings as opposed to the lesser influence of rather taciturn participants. The voting ballots were processed by the ranked-pair algorithm which determines the ordered list of winners of this "voting process". A quote ('") entry in the rank column indicates a tie: the corresponding word has the same rank as the previous word in the column.*

Rejection Experiment				Prohibition Experiment			
All Words		Salient Words		All Words		Salient Words	
Rank	Word	Rank	Word	Rank	Word	Rank	Word
1	you	1	triangle	1	you	1	square
2	the	2	no	2	the	2	no
3	a	"	square	3	this	3	circle
4	like	3	heart	4	a	4	moon
"	no	4	moon	5	is	5	triangle
5	this	5	circle	6	heart	6	heart
6	square	6	yes	7	one	"	one
7	one	7	like	8	to	7	this
8	that	8	it	9	no	8	yes
9	do	9	squares	10	like	9	like
10	moon	10	this	11	triangle	"	ok
11	it	11	one	"	square	10	again
12	triangle	12	ok	12	it	11	shape
13	circle	"	again	13	it's	12	it
14	it's	13	circles	14	that	13	crescent
15	heart	14	good	"	and	14	you
16	to	15	oh	15	circle	15	good
17	yes	16	right	16	very	16	very
18	is	17	triangles	17	do	17	ok(2)
19	ok	18	that	18	moon	18	right
20	oh	19	Deechee(2)	19	that's	19	circles
21	want	20	you	20	shape	20	round
22	don't	21	about	21	can	21	Deechee(2)
23	well	22	done	22	we	22	done
24	circles	23	ah	23	at	23	today

Rejection Corpus Salient Words (RCS): Ranks by vote vs. Ranks by frequency Again, when comparing the voting- with the frequency-based ranking no exceptional differences stand out. *Circles* moved from rank 7 to 13 and *this* moved up to rank 10. *Triangle* and *square* swapped ranks, but all object labels in their singular word forms are still part of the top 10 and so is *no*. Extending the scope to the top 20 the only remarkable observation is that the first phonetic form of *Deechee* disappeared from the listed 23 ranks - it has moved to rank 31. All other changes are minor rank swaps within the words that have already been part of the group.

Prohibition Corpus (PC): Ranks by vote vs. Ranks by frequency Considering only the top 10 in the two rankings, the only mentionable difference is the fact that, in the vote-based ranking *heart* moved from rank 11 up to rank 6 and, in exchange, *it's* moved out of the top 10 to rank 13. All other changes are minor rank-swaps amongst the words that were already part of the top 10. Equally in the top 20, no remarkable changes are discernible, the only additional word entering the top 20 being *very* - it's on rank 24 on the frequency based list.

Prohibition Corpus Salient Words (PCS): Ranks by vote vs. Ranks by frequency Apart from *yes* moving up by 3 to rank 8 and *again* moving from rank 15 to rank 10 and some minor rank swaps nothing notable can be said when comparing the top 10 vote- versus the frequency-based ranking table. When extending the scope to the ranks 11 to 20 the only mentionable observation might be the rank-increase of *you* by 7 to rank 14. Two new words made it into the top 20, the first of which is *round*, on rank 20, and which is in the frequency-based table on rank 27, and the second of which is *right* on rank 18 which ranks on the frequency-based list on 26.

In summary we can say, that the differences between the frequency- and the vote-based

rankings are rather miniscule. The top-ranking words, at least the top 20, seem to be indeed representative of the frequency-based rankings of single participants as far as such accumulative lists can be. There is no indication that the frequency-based rankings would be biased to such a degree that words would be amongst the top 20, that do not "deserve" to be there or that the rankings would be skewed through over-representation of the most communicative participants' frequencies.

5.2.3 Comparison with Saunders et al. (2012)

In this section we will compare the two corpora resulting from the negation experiments with the one originating from Saunders' experiment. Table 5.29 shows the ranks and frequencies of selected words similar to those shown for the prohibition and rejection corpora. It is important to note that the objects in Saunders' experiment were coloured and the objects sketches printed on the boxes varied in size. For this reason we find colour words and adjectives describing size in the corpus. Table 5.30 shows ranks and frequencies for salient words only. We will henceforth refer to these two corpora as *Saunders' corpus (SC)* and *Saunders' salient words corpus (SCS)*. When comparing the RC and PC with Saunders' corpus the following differences are salient:

- *You* slipped down in the ranks considerably: within the *SC* it ranks 10th and its coverage decreased by nearly two thirds (2.41%)
- *Like* also dropped very considerably: from the ranks 3 and 7 in the *RC* and *PC* respectively to rank 31 in the SC. Its coverage dropped from 3.2% (*PC*) and 3.31% (*RC*) to 0.81% (*SC*), a drop of nearly 75% relative to the other two coverages.
- *No* dropped from the ranks 6 (*RC*) and 9 (*PC*) to 65 within the *SC*. The coverage dropped from from 2.39% (*RC*) and 1.79% (*PC*) to 0.35% (*SC*). This is an even

Table 5.29: ***Word-frequencies of all words in the experiment of (Saunders et al. 2012).*** *Listed are the ten most frequent words within said experiment across all participants and sessions. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words uttered during the entire experiment. Apart from the highest-ranking words the same statistics are given for object labels, object properties, negation words, and words linked to the motivational state of the robot. See table B.9 in the appendix for the complete listing.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	a	702	8.71	*(23)*	shape	88	1.09	*(73)*	done	22	0.27
(2)	this	367	4.55	*(24)*	right	87	1.08	*(75)*	don't	21	0.26
(3)	blue	347	4.31	*(25)*	box	87	1.08	*(80)*	crescent	17	0.21
(4)	is	322	4	*(28)*	small	76	0.94	*(101)*	not	14	0.17
(5)	and	314	3.9	*(30)*	square	69	0.86	*(107)*	colours	11	0.14
(6)	red	302	3.75	*(31)*	like	65	0.81	*(109)*	isn't	11	0.14
(7)	green	286	3.55	*(32)*	star	61	0.76	*(137)*	nice	6	0.07
(8)	the	265	3.29	*(40)*	bigger	41	0.51	*(150)*	can't	4	0.05
(9)	that's	237	2.94	*(41)*	white	41	0.51	*(162)*	didn't	3	0.04
(10)	you	194	2.41	*(42)*	large	40	0.5	*(165)*	favourite	3	0.04
(11)	it's	161	2	*(44)*	colour	37	0.46	*(176)*	aren't	3	0.04
(12)	heart	160	1.99	*(53)*	Deechee	33	0.41	*(177)*	squares	3	0.04
(13)	circle	149	1.85	*(54)*	yes	33	0.41	*(178)*	circles	3	0.04
(14)	arrow	148	1.84	*(64)*	smile	28	0.35	*(229)*	triangle	1	0.01
(15)	side	146	1.81	*(65)*	no	28	0.35	*(260)*	never	1	0.01
(16)	cross	120	1.49	*(68)*	shapes	25	0.31	*(264)*	happy	1	0.01
(20)	moon	95	1.18	*(72)*	big	22	0.27	*(273)*	excited	1	0.01

bigger decrease than the one observed with *like*, the decrease is more than 80%.

- Similarly to *RC* and *PC*, object labels in the grammatical singular are mainly on the ranks 10 to 20. One is "pushed" beyond the 20, most probably because adjectives referring to colours and sizes take up some of the top 20 ranks.

- *Don't* similarly to *no* suffered a rather extreme decrease: in the *SC* it ranks 75th and covers 0.26%. This is a 70% drop compared to its coverage within *PC* and even more compared to *RC*.

- *Deechee*, the vocative, also suffers a considerable decrease down to rank 53. Its

coverage more than halves compared to the *PC*, the ratio is much worse for the *RC*.

- We see colour words entering the top 20 and even the top 10: 3 out of 10 words in the top ten are colour words.

Table 5.30: ***Word-frequencies of salient words in the experiment of (Saunders et al. 2012)****. Listed are the ten most salient words within said experiment across all participants and sessions. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words uttered during the entire experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot. See table B.10 in the appendix for the complete listing.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	blue	157	6.91	*(19)*	right	35	1.54	*(49)*	no	10	0.44
(2)	red	126	5.54	*(23)*	colour	24	1.06	*(50)*	it's	10	0.44
(3)	circle	117	5.15	*(24)*	good	22	0.97	*(52)*	done	10	0.44
(4)	heart	108	4.75	*(25)*	bigger	22	0.97	*(54)*	the	8	0.35
(5)	green	99	4.36	*(26)*	a	21	0.92	*(67)*	don't	6	0.26
(6)	arrow	81	3.56	*(27)*	that's	20	0.88	*(71)*	isn't	6	0.26
(7)	cross	79	3.48	*(28)*	Deechee	19	0.84	*(72)*	big	6	0.26
(8)	side	79	3.48	*(29)*	shapes	18	0.79	*(78)*	colours	5	0.22
(9)	box	64	2.82	*(30)*	it	18	0.79	*(91)*	didn't	3	0.13
(10)	shape	55	2.42	*(32)*	you	18	0.79	*(93)*	not	3	0.13
(11)	and	48	2.11	*(33)*	smile	17	0.75	*(97)*	favourite	3	0.13
(12)	moon	48	2.11	*(36)*	white	17	0.75	*(100)*	circles	3	0.13
(13)	square	47	2.07	*(38)*	large	16	0.7	*(107)*	nice	3	0.13
(14)	this	46	2.02	*(41)*	yes	13	0.57	*(119)*	squares	2	0.09
(15)	star	42	1.85	*(43)*	crescent	13	0.57	*(146)*	can't	1	0.04
(17)	is	40	1.76	*(46)*	like	11	0.48	*(156)*	aren't	1	0.04
(18)	small	35	1.54	*(47)*	yea	11	0.48				

Prosodically Salient Words When comparing the ranks and frequencies of all words (table 5.29) with the ranks and frequencies of only salient words (table 5.30) we can observe that the trend for function words to disappear from the top ranks is even stronger in Saunders' corpus. The top 10 of the SCS does not contain a single function word and consists entirely of object labels and other object-related words such as colours and more abstract words referring to the objects or parts of them such as *box* or *side*. The function words *a* and *the* moved down to the ranks 26 and 54 with salience rates slightly higher but still comparable to those found in the *RCS* and *PCS*. *A* and *the* have salience rates of 2.99% and 3.02% respectively compared to combined salience rates of 1.92% and 2.78% in the *RC(S)*

Table 5.31: *Prosodic salience rates of selected word groups and words. Saunders' Corpus - object labels:* heart, circle, arrow, cross, moon, square, star, crescent, triangle, *and their plurals; emotion words:* happy, smile, excited*; demonstratives:* this, that*; pers. pronouns:* you, we, I*; articles:* a, the*; vocative:* Deechee

word (group)	saliency rate (%)
object labels	65.20
emotion words	60
vocative	57.58
yes	39.39
no	35.71
don't	28.57
one	28.72
can't	25
it	24
like	16.92
demonstratives	13.82
to	11.43
pers. pronouns	5.69
articles	3.00
do	0

and *PC(S)* respectively (cf. table 5.31). When comparing the salience rates of the *RC(S)* and *PC(S)* corpora with the salience rate of the *SC(S)* corpora most of these rates seem to be generally lower in the *SC(S)* with few exceptions. The exceptions are articles and personal prononouns which have marginally higher salience rates in the *SC(S)*. Interestingly, also emotion words have a higher salience rate in the *SC(S)* than the *PC(S)* (60% vs. 54.90%) but still a lower salience rate than emotion words in the *RC(S)* (71.28%). Also *don't* has a higher salience rate in the *SC(S)* (28.57%) than in the *RC(S)* (23.79%) and *PC(S)* (19.65%).

Negation Words - Comparison between *RC(S)*, *PC(S)*, and *SC(S)* When we compare the "major players" amongst negation words in terms of frequency from the *RC(S)* and *PC(S)* to Saunders' corpora we find that their general frequencies are much lower in the latter corpora. *No* has the frequencies of 417 within the *RC*, 461 within the *PC*, but only 28 within the *SC*. Even if we account for the fact that the per session duration underlying *RC* and *PC* it approximately twice as long compared to *SC* and there is one participant less in Saunders' experiment, the difference is still remarkable.

If participants with Saunders' experiment had produced approximately the same number of *no*'s as in the negation experiments we would expect the frequency to be somewhere around 180[25]. Yet, as we can see, it is only about 15% of this 'target value'. This observation does not come as a total surprise as the utterance-level analysis already indicated that participants within Saunders' experiment produced markedly less negative utterances as compared to the negation experiments.

What was not already indicated by the utterance-level analysis is the circumstance that the salience rate of *no* is markedly lower than the latter rate within the negation corpora. This means that participants in Saunders' experiment not only produced markedly less *no*'s, but when they did so, they markedly less often prosodically emphasised the word. The latter cannot be said about *don't*. *Don't* reaches for reasons which we cannot explain on this analytical level the highest salience rates within Saunders' corpora (28.57% compared to 23.79% (*RC(S)*) and 19.65% *PC(S)*)[26] The overall frequency of *don't* is markedly lower in the *SC* (21) as compared to both the *RC* (269[27]) and *PC* (229). Again, the shorter session duration, factor 0.5, cannot account for this large difference, as we would expect a target value of around 100 if a similar production rate of *don't* would pertain. Regarding the production rates of *not* the findings are similar. Within the *SC* *not* was produced 14 times, whereas the frequencies within *RC* and *PC* are 118 and 198 respectively. Even if we take the lower value, 118, as reference, we would expect a target value of around 50 within *SC* if the production rate was similar. This is clearly not the case. The salience rates of *not* are 21.43%, 20.34%, and 15.66% within *SC(S)*, *RC(S)*, and *PC(S)* respectively

[25]We apply a scaling factor of 0.5*0.9 (50% duration time times 90% of number of participants).

[26]In order to explain why this is the case we would have to look at the pragmatic level and investigate the video recordings in order to determine in which situational context *don't* was produced and, furthermore, what communicative function participants 'had in mind' when doing so. For the negation corpora this is precisely what we did and the results of which will be reported on in the next section. Due to time constraints, we are not able to do this for Saunders' corpora within this thesis, which is admittedly somewhat unsatisfying.

[27]Combined frequency of both phonetic variants

and therefore seem to be roughly comparable.

Summarily we can conclude that within the *SC(S)*, despite comparatively high salience rates of two of the three most frequent negation words, the overall low frequencies of these words cause them to play only minor roles within the *SCS*. As the corpora of salient words form the basis of the robots active lexicon, negation words are very unlikely to be produced by the robot in Saunders' experiment. In strong contrast to this, negation words are very frequent within the negation exerpiments and were produced with similar or, in the case of *no*, markedly higher prosodic salience. Subsequently these words rank very high in the salient word corpora of these experiments and are therefore likely to be produced by the robot.

Interpersonal Function: Comparison between negation corpora and Saunders' corpus All percentages given here are percentages relative to the number of utterances in each corpus[28], not the coverage (which is a percentage relative to the number of words in the corpus). *You* yields within *SC* a percentage of 8.53% as opposed to 24.13% (*RC*) and 21.60% (*PC*), thus a considerable decrease. The percentage of use of the vocative within *SC* amount to 1.45% compared to 4.33% (*RC*) and 3.16% (*PC*). 5.49% of all utterances contain a *we* or a variant[29] compared to 3.92% (*RC*) and 4.49% (*PC*), thus a slight rise. Finally, *I* and variants[30] occur in 1.8% of the utterances of *SC*, compared to 3.33% within *RC* and 3.13 % within *PC*.

The percentages of the above stated indicators of interpersonal function within *SC* ex-

[28] Assuming that each of the respective words occur at most once each in an utterance, allows us to make statements such as "23% of all utterances contain a X.". We know that this assumption does not hold for every utterance and word. "You know that you can't have that" is a counter-example. But based on our experience when coding the language used by participants we believe that it is approximately true and it allows us to speak about word-utterance relations in a more intuitive manner.

[29] *We* has only one variant within *SC* which is *we've*.

[30] Variants of *I* within *SC*: *I'm*, *I'll*

cluding *you* sum up to 8.74% compared to 11.58% (RC) and 10.78% (PC). The difference of 2.04 constitutes a 18.92% decrease compared to the 10.78% level of PC. Without statistical testing no decision can be made if this decrease is significant, thus we can only state that the difference in terms of these indicators between the negation corpora and Saunders' corpus is nearly three times the difference (factor 2.74) between the both negation corpora. Furthermore the considerably lower rate in the use of *you* which decreased by more than 60% seems striking but due to the reasons given in section 5.2 we hesitate to take this as an indicator for interpersonal function.

Emotion/Volition words: Comparison between negation corpora and Saunders' corpus This paragraph is intended to shed some light on the differences between the corpora in terms of emotion words. By analogy with the analysis of interpersonal function we compare in this section the percentage of the emotion words *smile*, *sad*, and *happy* relative to the number of utterances of the corresponding corpus. Another word of interest with a relatively high rank in the negation corpora in this context is *like*. It is mainly used by participants to ask Deechee if it likes something, i.e. if it wants to talk about a particular object. Thus it is often, probably in the majority of cases, used within utterances which we will later call *motivation-dependent questions*. It is unclear to the author if these questions really refer to the emotions of the recipient or if they refer to the will of the recipient (emotion vs. volition). As the question "Do you like X" can be replaced with "Do you want X" we are inclined to count it as a question involving volition rather than emotion[31].

[31]From a conversation-analytical standpoint one could even argue that it does not matter to what precisely "like" refers and that the obsession with identifying the 'correct' referent is misleading. We do know precisely what kind of conversational work a question such as "Do you want an ice cream" in an appropriate context accomplishes. We also know that the precise referent does not matter to the agent that performs it. The 'work' that is performed by the utterance is independent of the issue whether the recipient really 'wants' (volition) an ice cream or if it 'just feels like' one (emotion). The forced quest to

The percentage of *smile* within *SC* is 1.23% compared to 0.24% and 0.28% within *RC* and *PC* respectively. *Sad* was never produced with *SC* at all whereas its percentages within *RC* and *PC* are 0.81% and 0.11% respectively. *Happy* occurs in 0.04% of all utterances of *SC* as compared to 0.72% in the case of *RC* and 0.30% in the case of *PC*. *Like* finally, the word whose membership to the class of emotion words is dubious at least, shows the biggest differences across corpora. It is used in 2.86% of all utterances of *SC* compared to 10.91% and 7.09% within *RC* and *PC* respectively.

If we sum up these percentages, excluding *like*, the picture is mixed at best: Within Saunders' corpus 1.27% of all utterances contain emotion words compared to 1.77% of the *Rejection Corpus* and 0.69% of the *Prohibition Corpus*, putting *Saunders' Corpus* in between the two other corpora. If *like* is counted as an emotion word and included in this sum the picture changes dramatically leading to 4.13% within the *SC* as compared to 12.68% and 7.78% within the *RC* and *PC* respectively. As we are not readily willing to count *like* in the class of emotion words we cannot assert that the negation corpora would contain more emotion words per utterance than Saunders' corpus. By the same token the difference in the use of *like* between the negation corpora and Saunders corpus is very marked. If we deny its membership to the class of emotion words, we can't see any other option than to count it in the class of volition/intention words such as *want*. If we accept this classification the marked difference in the use of *like* indicates a marked difference in the participant's perception of the robots intentionality. If there is indeed a marked difference in the perception of intentionality we would expect a similar discrepancy with other volition words. And indeed, when looking at the production frequencies of *want* plus variant[32] we find such a marked difference: Within *Saunders' Corpus* only 9 *want*'s

identify a referent for each and every content word in an utterance might be just another example of what happens when "language idles"(Wittgenstein 1984).

[32]There is only one variant of *want* within the three corpora: *wanna*.

were produced which correspond to 0.4% of all utterances. Within the *RC* and *PC* 3.54% and 3.24% of all utterances contain a *want* or a variant thereof. Thus the difference in the production rates of *want* and variant is even more marked than is the case with *like*. The negation corpora have more than 8 times as many *want*'s and variant per utterance than Saunders' corpus.

5.3 Human + Robot: Pragmatic Level

In the previous two sections we have established that in both rejection as prohibition experiment a high number of negative utterances were produced by the participants of the respective experiments, when compared to a very similar noun-learning experiment, Saunders' experiment. The reason for this increase in production was hypothesized to be the presence of emotional displays via facial expressions and motivationally congruent body behaviours.

The reason that we still can only hypothesize instead of prove that this is the case is the presence of three potentially confounding variables between the experiments under comparison. We named three potentially confounding variables: differing length of the sessions, the fact that the objects in Saunders' experiments were coloured and varied in size, and slightly differing instructions to the participants which made them aware that the robot would have preferences towards the objects. The author deems it highly unlikely that the first two of these variables, session length and differing object colour and size would trigger a difference in the linguistic behaviour manifested in an marked increase of the use of negation. We only consider the third potentially confounding variable, the hinting of participants towards the robot having likes and dislikes, a potential suspect for influencing the participants behaviour. Intuitively the effect of this potential case of priming would seem to be an increase of emotional and volitional words produced by the participants. In point of fact the word-level analysis supports this intuition. We do observe a large increase of volitional utterances containing words such as *like* and *want*. Yet it is impossible to tell how much of this increase is caused by potential priming and how much of it is caused by the robot actually exhibiting behaviour that indicates the presence of volitional and motivational states. Nevertheless the assumption of this kind of priming

having taken place in conjunction with the observation of a similarly strong increase in negative utterances turns out support an hypothesis that is very much akin to our major hypothesis: Motivational and negative utterances are heavily interrelated.

The heightened number of negative utterances in the negation experiments were subsequently shown to be mirrored by higher frequencies of negative words on the word level. Moreover, the analysis of the word corpora revealed that the majority of negation words that were either used within those negative utterances or that constitute single-word utterances in their own right pertained to a very small set of word forms. By far the most frequently used negation word turned out to be *no*, followed by *don't*, and, with a considerably lower frequency, *can't*. In case of the prohibition experiment a fourth word, *can't*, was shown to have been produced frequently.

A question that both of these analyses cannot answer, due to their focus on sets of words or sets of utterances, is the question to what purpose participants produced these utterances. This question is closely related to the question for the communicative functions of these utterances. Communicative functions pertain to the pragmatic level of linguistic theory and cannot be reduced to word corpora or grammatical phenomena (cf. section 2.1.2).

In order to determine the kinds of communicative functions associated with the negative utterances produced by participants within the negation experiments we decided to construct two taxonomies both of which are based on the taxonomy for negative utterances developed by Roy D. Pea which was summarized in section 2.5.

5.3.1 Overview of section

Subsection 5.3.2 describes the process of how the negation taxonomies for human and robot utterances were constructed.

Subsection 5.3.4 then presents the taxonomy of negation types for human negation, which our participants engaged in, whereas subsection 5.3.3 presents the taxonomy of robot negation types, i.e. the kinds of negation which the robot engaged in during the experiments. Both taxonomies are the outcome of said construction procedure.

Subsection 5.3.5 subsequently describes the evaluation of both taxonomies by means of intercoder agreement. This evaluation was performed in order to ensure that the taxonomies and quantitative results based on these taxonomies can be replicated by coders other than the author. A sufficient degree of intercoder agreement ensures a certain level of reproducibility if the taxonomy is applied by other researchers to the given or similar conversational data. We do not claim that the given taxonomies are general in the sense that they would sufficiently cover negative utterances produced within scenarios or forms of life that differ considerably from our experimental scenarios. We rather suspect that many more negation types would have to be introduced into both taxonomies in order to account for different situational contexts before any potential claim for generality could be made.

As will be shown, the evaluation of the two taxonomies only yielded a good intercoder agreement for the human negation taxonomy but a moderate agreement for the taxonomy of robot negation. We therefore attempted to improve the robot taxonomy by two different procedures.

In subsection 5.3.6 we present an automatic optimization approach based on Cohen's κ, a coefficient which numerically quantifies the amount of agreement, and report the results of this automatic optimization.

In subsection 5.3.7 an account of a complementary qualitative optimization attempt will be given which is based on a structured interview. During this interview we determined the reasons for the particular decisions of the two coders specifically for those codes

in the,coding table where the two disagreed from each other. Interestingly the results from this two methodologically very different methods yield comparable results which will be reflected upon in this subsection. The results hint towards potential limitations with regards to the application of conversation analysis, and derivative methods, to conversations between strongly asymmetrical conversation partners, i.e. conversation partners with significant differences in their respective communicative competence.

5.3.2 Construction of Taxonomies from the Utterance Corpora

In order to assess the number and kinds of communicative functions associated with the negative utterances encountered in our experiment we constructed two taxonomies: one for the negative utterances produced by the participants and a second one for the negative utterances eventually produced by the robot. As a starting point for the construction served the taxonomy of negation in early child language proposed by Pea (1980). Our intention behind adopting Pea's taxonomy as template for our taxonomies is to ensure that our results would be at least partially comparable to potential results obtained by research of negation in early child language and to render our results as informative as possible for research on human language development. The taxonomies were constructed by the author who also acted as the first coder.

Construction of the Robot's Negation Taxonomy The first taxonomy to be constructed was the taxonomy of negative robot utterances. To this purpose all of the robot's negative utterances were extracted from the corpus together with their respective session ids and time stamps. For each utterance the author subsequently looked at the video-taped conversation in the course of which the utterance occurred. Contingent on the observed situational and conversational context we attempted to match each utterance to one of

Pea's negation types. If none of these was deemed appropriate, a new type was created. After having processed all of the robot's negative utterances in this way, we looked again at those particular utterances which were classified as members of negation types that had only few member utterances in order to determine if some of these types could be eliminated by merging them with other types. This attempt to eliminate rarely occurring negation types was performed in order to keep the taxonomy minimal. The taxonomy that resulted from this procedure is depicted in figure 5.4.

Construction of the Participants' Negation Taxonomy Pea's taxonomy, originally constructed to classify negative utterances of toddlers, incorporates the notion of adjacency on the topmost level (cf. figure 2.2). Due to this property his taxonomy lends itself as a starting point for constructing a partially symmetrical taxonomy for the negative utterances of the parent, participant, or, more generally, the linguistically more competent conversation partner. We observed, and so it seems did Pea, that many negative utterances are not produced in isolation but rather occur as one part of an adjacency pair. This is not specific to negative utterances but a well documented structural property of many human face-to-face conversations (cf. section 2.1.3). As it will be shown later in this section participants often produced negative first pair-parts, typically negative questions such as "You don't like it?" that make relevant another, potentially negative, second pair-part such as a simple "No!".

Thus, looking out for potential negative first pair-part negation types that fit the negative second pair-parts of the robot's taxonomy (i.e. the left side of the classification tree in figure 5.4), the construction of the participants' negation taxonomy followed the same procedure as the one in place when constructing the robot's taxonomy. All negative utterances of all participants and all sessions were automatically extracted from the corpus

together with their respective session ids and time stamps. As with the robot's utterances, all utterances were evaluated based on the conversational and situational context in which they were produced by watching the video recordings of the experimental sessions. Again, we tried to keep the taxonomy as minimal as possible, yet the resulting number of negation types is considerably higher as compared to the number of robot negation types. This circumstance is most probably due to the participant being the pro-active and dominant conversation partner who essentially leads the conversation by selecting, describing, and asking questions about particular objects. This asymmetrical conversational relationship can be expected between conversation partners with largely differing degrees of communicative competence and has also been observed with mother-child dyads (cf. section 2.1.3). The taxonomy resulting from this procedure is depicted in figure 5.5.

5.3.3 A Taxonomy for Robot Negation

Figure 5.4 visualizes the taxonomy of negation types the robot engaged in as judged by the author and derived from both experimental corpora. For a detailed description of each negation type with exemplary utterances we refer to paragraph B.8.4 of the coding scheme in the appendix.

Pragmatic Negation Words The original negative utterance set included utterances containing so called *pragmatic negatives*. These are still mentioned in the automatic optimization of the robot's negation types (subsection 5.3.6), and in the coding scheme in the appendix, but were removed from the current taxonomy after the evaluation of the latter.

The author observed during the first round of coding that the robot's production of some words which are typically not mentioned in the literature on negation can in particular situations have the same function or effect as words which would typically appear in lists of

lexical negation words ("no", "not", "gone", etc.) or which are uttered in the same situations in which participants often produced lexical negation words. The *pragmatic* but not *lexical negation words* on the robot's part were *go*, *oh*, *down*, or *done*, which at times seemed to have the same effect on participants as lexical negation words. On the participants' part *sad* was used by one participant in the same situations where other participants used negative intent interpretations such as "oh you don't like the heart". In the case of *go* and *oh*, due to the suboptimal performance of the speech synthesis, there were clear indications that they were sometimes misheard as *no* by some participants and could subsequently, by pure accident, have the same effect as an acoustically proper *no*. In the cases of *down*, or *done* there were no indications that they were ever misheard, yet these words have at times been interpreted in certain situations as rejective utterances with regards to some presented object. Initially we therefore selected not only proper lexical negation words for the pragmatic classification as specified in the coding scheme but also added utterances consisting of or containing these *pragmatic negatives* to the set of to be classified utterances. This was done in order to see how these words would fare compared to the 'proper', i.e. pragmatic *and* lexical negatives.

Upon completion of the classification it became evident that the intercoder agreement for the robot's negative utterances was particularly low due to a high degree of disagreement between coders with regard to these pragmatic negatives. We therefore decided to remove them from the final taxonomy. Yet they are still briefly mentioned in the coding scheme in the appendix as well as in subsection 5.3.5 which covers the evaluation of the taxonomies. Furthermore their presence during the classification procedure necessitated the presence of the *none*-type, which had to be introduced in order to cater for those cases in which a potential pragmatic negative did not "act" as a negative in particular utterances (cf. the coding scheme in subsection B.8.4 of the appendix).

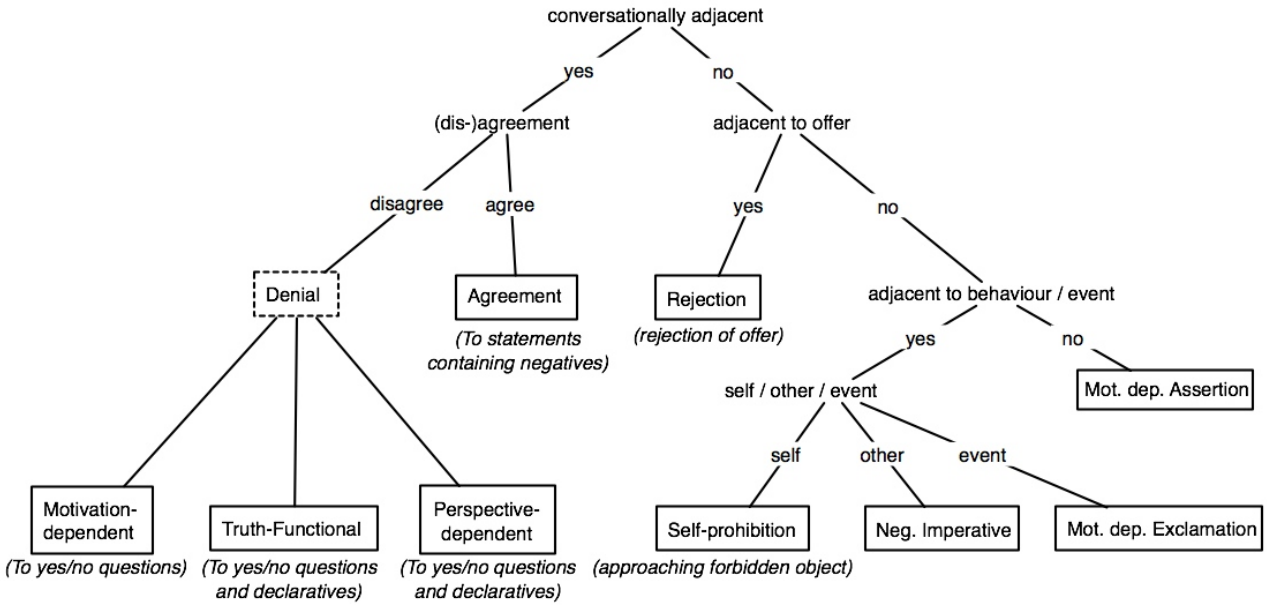


Figure 5.4: *Taxonomy of negation types used by robot within both experiments*

Comparison to Pea's taxonomy

Redefinition of particular negative meanings Importantly we extended or narrowed the definitions for the types *rejection*, *motivation-dependent denial*, and *perspective-dependent denial* compared to Pea's definitions for the same types. As outlined in section 2.5.2 Pea's definitions of and side remarks for *rejection* and *motivation-dependent denial* seem to indicate a certain overlap. We therefore narrowed the definition of *rejection* to include only negative responses to non-linguistic offers, that is we defined it to only refer to linguistically non-adjacent negation types. Conversely *motivation-dependent denial* was defined to denote all forms of negative robotic utterances that are linguistically adjacent to participants' utterances and which involve the motivational state of the robot. Furthermore *perspective-dependent denial* is defined in our taxonomy to explicitly include

negative responses that involve the physical perspective, the ability, or the knowledge of the robot. For each of these different kinds of dependencies, perspectival, ability-related, or epistemological, a separate negation type could have been created. Yet as there were only few occurrences of utterances deemed to be of these kinds we decided to capture all of these within a single category.

Lost negative meanings The following negation types or negative meanings, listed in Pea's taxonomy, are not present in our taxonomy: *disappearance* and *make-believe* on the adjacent and non-adjacent side, and *unfulfilled expectation* on the non-adjacent side. We did not consider any negative utterance from our corpus as falling into one of these categories.

New negative meanings Our taxonomy contains some new types which do not occur in Pea's taxonomy: *negative imperatives* are in some way similar to *rejections*. As we defined *rejections* rather narrowly to refer to rejections of non-linguistic offers, *negative imperatives* refer to all other kinds of 'rejective' negatives on part of the robot which are deemed by the coder to have been triggered by participants' behaviours other than offers.

Motivation-dependent assertions are a residual category to capture all negative utterances other than *rejections of offers* or *negative imperatives* which are non-adjacent and deemed to involve the motivational state of the robot. For an utterance to fall into this category the coder must judge it to be too weak either in terms of intonation or in terms of the situational context in order to count as *negative imperative*.

Finally we introduced *motivation-dependent exclamation* to cover both adjacent and non-adjacent negative utterances that are deemed to involve the motivational state of the robot, but unlike *negative imperatives* are deemed to refer to some event rather than an intentional action of the participant. An example would be a negative utterance that

accompanies or is triggered by the accidental drop of a glass. For further definitions of and examples for the various robot negation 'types' of the taxonomy we refer to section B.8.4 of the coding scheme in the appendix.

On the difference between the taxonomies An obvious question to ask when comparing Pea's taxonomy for negative child utterances with our taxonomy for negative robot utterances is why they differ from each other. The answer to this lies in the defining criteria for the various 'types'. Remember that all these 'types' or categories of negative meaning are defined via the conversational and situational context in which they occur (cf. section 2.5.2).

The two experimental scenarios in the context of which the robot produced these utterances, i.e. the situational context, differ in important ways from the context in which Pea's subjects produced their utterances. Probably the most important difference is that Pea's children were ambulatory, they could and did move around or were moved around by their mothers. Deechee on the other hand can't walk, can't move its body to another location. The robotic experiments were conducted in a single, small room. Participants did not go with Deechee into the kitchen, bathroom or any other room because the robot is attached to large, heavy, and immovable power supplies, and even it this was not the case, its sheer weight would prevent anybody from doing so.

In terms of the capacity of movement a human equivalent to Deechee would be a child that is paraplegic and has been so from birth, sits in a wheelchair, and yet worse, sits in a wheelchair that is nailed to the ground. Furthermore Deechee's capacity to manipulate objects is far worse than that of a one-year old toddler, and Deechee has no history of manipulating objects because of which neither participants nor coders have ever observed the robot doing so.

Another important difference between the robotic and the children's scenarios is the fact that the robot has no common history with any of the participants in a shared living environment, and for this very reason there are no entrenched habits that form part of the common ground between the two conversation partners.

It is for all of these reasons that the coders as external observers as well as the participants, both being perfectly aware of these circumstances, would never judge a negative produced by Deechee to be an example of unfulfilled expectations. Participants as coders know that the robot has no expectations of where to find the objects which are the topic of conversation, because they have never seen the robot crawling around and looking for them. The robot is physically unable to do so, therefore habitual locations are 'none of his business'.

Make-believe involves other, for robotic standards rather complicated bodily mechanisms such as "chuckling" and cheeky smiles, which Deechee is not capable of. Yet in some situations Deechee's smile came close to being interpreted as cheeky.

Disappearance negation was not observed for the simple reason that objects rarely disappeared, nor did participants make comments in this direction. Only when Deechee dropped an object was there the possibility of temporary disappearance. In the few cases where Deechee did produce utterances such as "oh", "no", or "done" immediately after having dropped a box, these were deemed by the first coder to be *motivation-dependent exclamations*, utterances expressing surprise or sorrow rather than comments on disappearance. This interpretation was lent support by one participant's reaction to such an utterance and/or event with a "No worries".

5.3.4 A Taxonomy for Human Negation

When looking at the taxonomy of negative meanings produced by the participants (cf. figure 5.5), it becomes immediately obvious that there are many more "types" or negative meanings as compared to the ones produced by the robot or as reported by Pea. We would like to argue that there are essentially two reasons why this is the case, which we will explicate in the following paragraphs. A second, visually salient difference in the participants' taxonomy graph is the additional distinction of conversational adjacency into first pair parts and second pair parts as well as the more explicit marking of types according to non-conversational adjacency. By analogy with Pea's use of *adjacency* our use of *first* and *second pair part* does not exclusively refer to parts of adjacency pairs in the strong conversation analytical sense but is used more liberally. It denotes the adjacency of two or more sequentially linked conversational turns across speakers but without the stronger requirement that the non-production of a second pair-part would lead to the second speaker being accountable for this non-production (cf. section 2.5.2).

The first reason why there are more than twice as many meanings of negation listed in the participants' taxonomy as compared to the robots' taxonomy is the lower degree of communicative competence on the robots' side. More precisely, the restrictions imposed by it being constrained to one-word utterances automatically excludes the possibility of it engaging in some negative meanings, which require more complex, grammatical utterances.

What we termed *quoted negation* for example, requires the to-be-quoted or reported negative to be embedded in a main clause as in "I said "no" ". If there is no embedding main clause, there can not be any embedded quote. One could argue that this meaning 'type' is none, and that the communicative function captured by this form of negation might be assimilated into other 'types'.

Negative tag questions are possibly the most grammatically defined type here and are

for this very reason beyond the productive scope of a child (or robot) in the one-word stage. An example of an utterances falling into the category of tag questions is "You do like the square, don't you?", with the negative tag question being the terminal clause after the comma. One might argue that these negative utterances are not 'real' or semantic negatives at all and should be excluded from the taxonomy. Indeed negative tag questions are at least a borderline case.

A second reason for the lesser number of categories of negative meanings, a reason which applies to the robot's situation but also to the toddlers in Pea's study, is the asymmetrical communicative relation between the two interlocutors, mother and child or participant and robot, where the one with the higher communicative competence typically takes the conversational lead, for example by proactively asking questions and thereby constructing a conversational 'scaffold' and conversational slots for the conversationally weaker partner. Our robot, and judging by Pea's taxonomy, 'his' toddlers appear not to produce questions containing negatives. More generally the robot does not produce questions at all. Similarly there are indications from research on parent-child dialogues during the one-word stage that questions are very prevalent within the parents' productions but very scarce within the toddlers' productions (cf. section 2.2.2). The presence of 'proper' negative questions, i.e. first part-pairs of *question-answer* adjacency pairs, in the participants' speech accounts for four types in the participants' taxonomy which are absent in both the robot's as in the toddler's taxonomies.

Apart from the communicative asymmetry, the lack of grammatical or lexical competence required to produce negative questions may account for a lack of these in the toddlers' productions. As opposed to their positive counterparts questions containing negatives are typically skewed in terms of neutrality. This is to say that the participants' positive questions were typically neutral in terms of the the expected response. For example the question

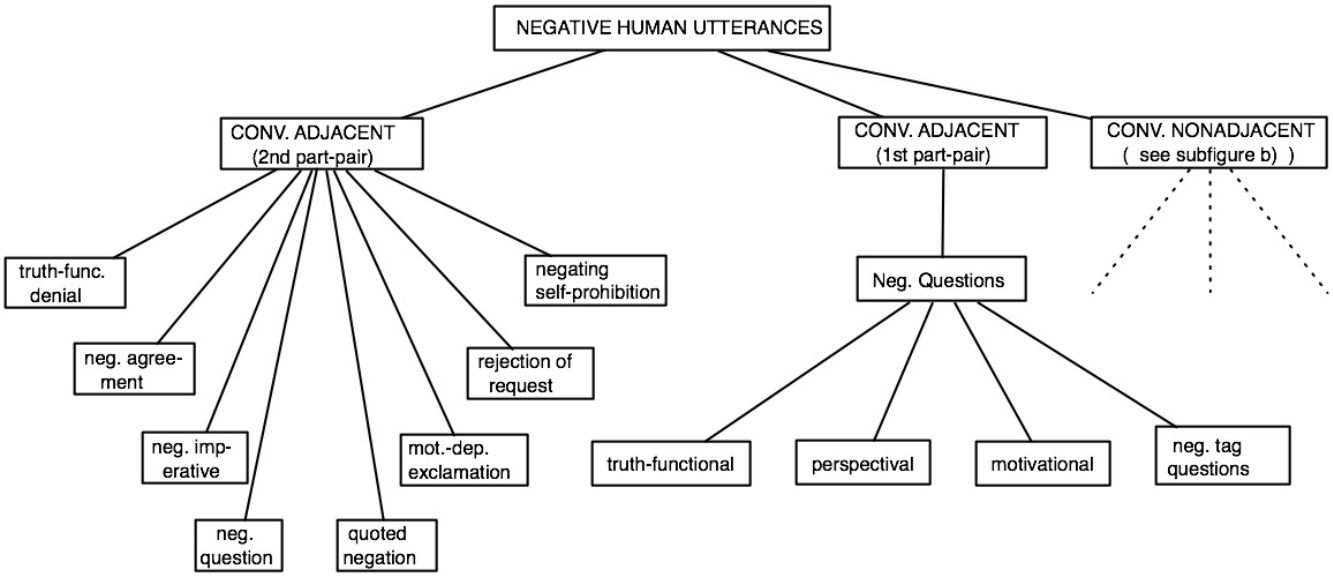
"Do you like Marmite?" does not indicate any bias in the speakers expectation towards the answer, both "yes" and "no" appear to be expected with equal probability. In contrast the speaker's expectation in the negative variant "Don't you like Marmite?" is biased towards a negative response based on the speakers experience with the interlocutor such as the observation of the addressee producing a negative facial expression upon biting into a Marmite-laced sandwich or the like. In this regard negative questions convey more information about the speaker's expectation than their positive counterparts. This additional information is then evidently expressed with an additional negative construction, here the postfix "n't" or additional negative words[33]. Hence if a communicatively 'weak' speaker such as a toddler or our robot is not able to produce the three elements required by a negative question, 'negator' + question particle/ interrogative intonation contour/ gesture + topic indicator, such questions can neither be produced nor identified as such by the participant or an observer. For our robot we certainly know that it cannot produce the required intonation contour commonly used with questions nor was it programmed to use gestures indicating negation or question.

In this context we should note that the coders found it often difficult on the participants' side to distinguish between *negative motivational questions* and *negative intent interpretations*. This difficulty arose because participants often did not leave sufficient pauses after the supposed *negative motivational questions*. This subsequently raised the coders' doubt that these were indeed meant as 'proper' questions, i.e. questions where the speaker expects and therefore waits for an answer. Yet the respective utterances often had the intonation contour of a question as opposed to the 'assertive contour' of *negative intent interpretations*.

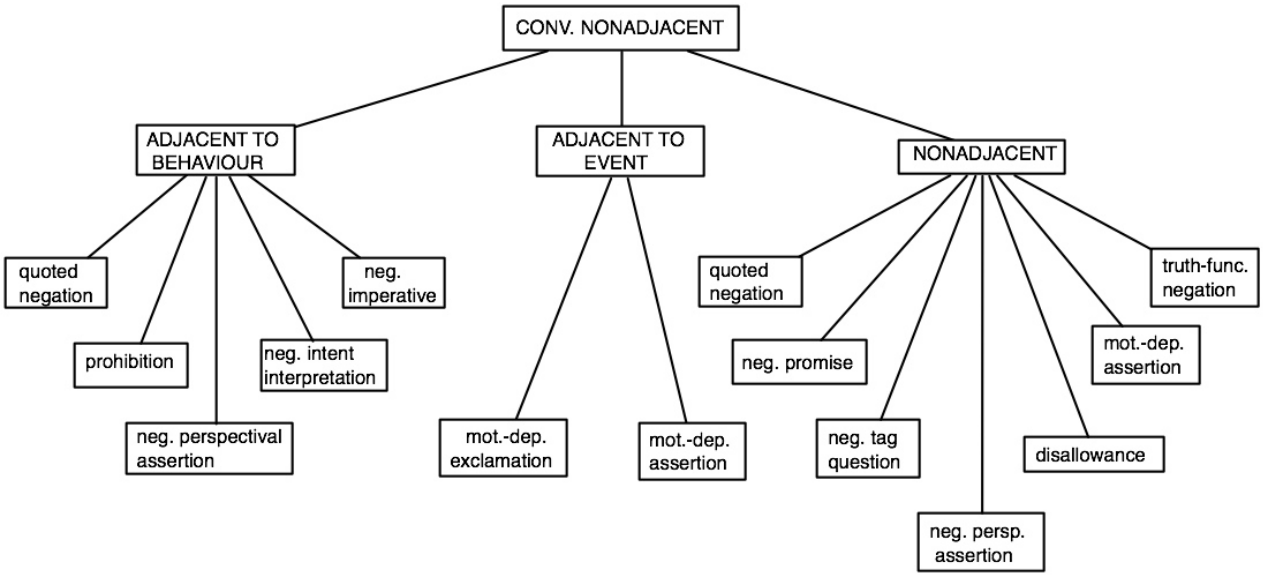
This is one example that shows that the treatment of utterances in isolation may

[33]If we imagine pidgin variants of this question such as "like Marmite?" vs. "no like Marmite?" the negative variant requires one word more and is therefore lexically more complex rather than grammatically.

evoke a false certainty in the observer that these could be unambiguously classified by just paying enough attention to the details of production such as intonation, grammatical indicators like word order, etc. Yet as soon as we take into account the conversational context, with gap lengths between sequential turns being an important property of the latter, the presumed, theory-driven 'ease of classification' shows severe incompatibilities with the actual "language at work".



(a) *Human Negation Types pt. 1*



(b) *Human Negation Types pt. 2*

Figure 5.5: *Taxonomy of negation types used by participants within both experimental scenarios. Conv.: conversationally, 1st part-pair, 2nd part pair: parts of an adjacency pair such as question (1st part-pair) - answer (2nd part-pair)*

Notice that we also introduced the 2nd part-pair question 'type' *negative question* into the participants' taxonomy. These are questions mainly by intonational contour and were judged by us not to be questions for semantic information but rather questions regarding the sincerity of the robot's utterance. Thus, if lexically expanded, these questions would amount to something like "Is that really what you mean?" or "Are you serious?", but are in the recorded conversations typically expressed by a simple "no?" (cf. section B.8.5 in the coding scheme). Utterances of this kind show a certain smiliarity to *negative agreements* in that the robot's negative word is repeated by the participant. Yet the intonation contour of *negative agreements* indicate certainty on the part of the participant with regards to the communicative intent of the robot, whereas *negative questions* express the participants' doubt.

Other additional negation 'types' which are typically not expressed with a simple one-word utterance are *negative promises* and weaker forms of future commitments such as "I won't do X again" or "I promise not to do X". Utterances of this kind typically involve a word or construction which expresses future commitment such as "promise", or "won't", and one or more words that specify the referent of what the speaker commits herself to such as "(not) doing X". Theoretically such a promise could be performed in one word if the referent of the promise is clear or salient to both conversation partners at the time of production. For example a conversationally weak speaker B, could commit herself not to do something in the future by a simple "won't", if a conversationally strong speaker had specified the referent adjacently in a first pair-part such as "Promise me not to do X any more". Yet we never encountered utterances produced by the robot that seemed to fit this pattern. The absence of an equivalent type in Pea's taxonomy indicates that toddlers in the one-word stage typically also do not commit to certain future behaviours in a linguistic

manner.

For a description of the other 'types' of negative meaning and examples for each 'type' we refer to section B.8.5 of the coding scheme.

5.3.5 Evaluation of the Taxonomies

A potential problem of 'hand-crafted' taxonomies that rely on the subjective assessment of their constructor is a certain degree of arbitrariness. On the pragmatic or functional level of human communication one may even expect a certain degree of indeterminacy. This is indicated by the well-documented presence of conversational repair mechanisms which are employed whenever one conversational partners misinterprets one or more utterances of the other. These repair mechanisms would not be needed if every conversational move could be unambiguously 'decoded' and correctly understood by the interlocutor within a given conversation. Both, the fact that conversational repair mechanism are one of the best documented conversational phenomena, and the circumstance that participants of a conversation display to each other how they understood each other's utterances, are strong indicators that single utterances are often inherently ambiguous. This is even the case when these utterances are embedded in a particular conversational context.

The coders qua external observers of a conversation and members of the same language community rely on the same conversational resources as the participants in order to determine the function or meaning of a particular utterance. We therefore cannot envision any method to radically reduce this ambiguity. Rather a certain degree of ambiguity seems to be part of the fabric of our language games. As language scientists and conversational observers we are therefore effectively limited in our understanding of these utterances by the same constraints as a conversationally fully competent participant. Subsequently all functional or pragmatic taxonomies which are derived from actual conversations must be

based on subjective evaluations and will therefore necessarily exhibit a certain degree of arbitrariness. Therefore the best we can do is to strive to minimise the degree of arbitrariness to an acceptable level, i.e. optimize the taxonomy by adding or removing types such that an acceptable level of agreement is reached.

To this purpose, we have to evaluate, i.e. quantify, the degree of arbitrariness or, inversely, the 'goodness' (Di Eugenio 2000) of our existing taxonomy. One method to do so is the introduction of one or more additional coders that perform the same classification task as the constructor, or first coder. The results can then subsequently be compared. Thus, in absence of a gold standard or 'god's truth' with regard to our pragmatic classes or types, we choose intercoder agreement as an inverse measure of the subjectivity or arbitrariness inherent within the taxonomy. High levels of agreement between the coders is then taken to be an indicator for a low degree of subjectivity if one accepts the definitions and examples of the taxonomy. The latter is described in a so called *coding scheme* which the additional coders are given as a manual in order to perform the classification. Which numeric levels of agreement are considered high or sufficient is a matter of discussion and tradition within the particular scientific community and subject area within which the coding task is performed. The actual numbers can differ quite considerably across different communities (cf. Di Eugenio 2000). A common method to quantify intercoder agreement is the calculation of Cohen's κ coefficient (Cohen 1960). This coefficient discounts chance agreement between coders which is not the case for simpler measurements such as the relative agreement between coders, i.e. the percentage of cases in which the coders agree in their decision relative to the total number of decisions (cf. Di Eugenio (2000) for a discussion of the use of Cohen's κ in language related coding tasks).

For our taxonomies the process of determining the intercoder agreement proceeded as follows: Upon completion of the taxonomy, which was constructed in parallel to the ini-

tial coding of all negative utterances, the author and first coder drafted a coding scheme which explains the various types of the two taxonomies and several other features which subsequently were to be coded by the second coder (cf. section B.8 in the appendix[34]). As second coder a lab assistant was chosen who previously was involved with the manual transcription and realignment of the audio recordings that were recorded during the experiments. The coding scheme was handed over to the second coder, 20% of negative utterances were randomly selected. Before coding the selected utterances for their type, the second coder was asked to decide and code for each utterance if she would deem it felicitous or adequate in the given situation. In a second stage the second coder then classified these utterances according to the coding scheme independently from her decisions on felicity[35]. It is important to note at this point, that the second coder was not trained on the data set prior to the coding proper[36]. Upon completion of the coding of the second coder, Cohen's κ was calculated as follows:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

Here $P(A)$ denotes the *observed agreement*, and $P(E)$ denotes the *expected agreement*. As the number of decisions in our coding task is finite, the calculation of these probabilities boils down to the calculation of ratios of frequencies. $P(A)$ is calculated by adding up the relative frequencies of each class or type where both coders agreed, i.e. by summing up the normalized entries on the main diagonal of the confusion matrix (cf. table 5.32, yet

[34]Note that many of the additional features specified in the coding scheme such as the participants' (non-)reaction to the robot's utterances were not used in the subsequent analysis and are therefore not included in this thesis

[35]For the details of how exactly the coding proceeded we refer to section B.8 of the appendix.

[36]An initial training of coders is sometimes performed, when several coders are used to code large data sets. In these cases the investigators obviously assume that their taxonomy is sufficiently good, and the purpose of stating the intercoder agreement is different from ours: there the purpose of the investigators in citing the intercoder agreement seems to be to ensure that their coders internalized the coding scheme sufficiently well and that the results are not skewed by a lax application of the scheme by the coders.

the entries there are absolute, non-normalized frequencies), where *normalized* is meant to mean *percentages relative to the total frequency count.* $P(E)$ is calculated by adding up the so called joint probabilities of the marginals (see Cohen 1960).

Gwet (2002) shows that Cohen's κ statistic tends to overestimate the expected or chance agreement depending on the marginal probabilities. This can lead to a distortion of the intercoder-agreement and Gwet therefore developed the alternative *AC1* statistic which corrects for this overestimation. For this reason we calculated additionally to the widespread κ statistic Gwet's *AC1* for both confusion matrices in order to ensure that our κ values are not grossly distorted. Yet, as most numerical boundaries that determine what can be regarded as an acceptable intercoder agreement are still expressed in terms of the κ statistic, we had to use the latter for orientation despite its statistical shortcomings.

Intercoder Agreement

Across all sessions and participants the robot produced 505 negative utterances including utterances containing or consisting of pragmatic negatives. 135 of these 505 utterances were subsequently coded by both coders out of which 37 (27.4%) were utterances with pragmatic negatives. Table 5.32 shows the confusion matrix for both coders' classifications of these negative utterances in terms of negation types. Also displayed are the corresponding κ and $AC1$ values. Table 5.33 shows the same statistics for negative utterances produced by the human participants.

Intercoder Agreement: Robot Negation Types The κ-value of 0.46 for the pragmatic classification of the robot's negative utterances is definitively on the very low end of all "agreed upon boundaries" for intercoder agreement. According to Di Eugenio (2000) this value is definitively too low on Krippendorff's (1980) scale as it is $< .67$, which is the

coder 1 \ coder 2	[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[0]	6	0	0	0	6	0	0	0	0
[1]	1	8	0	1	5	1	5	0	0
[2]	0	0	1	0	0	0	0	0	0
[3]	0	0	0	11	1	0	0	0	5
[4]	6	0	0	0	14	1	3	0	1
[5]	0	0	0	0	4	0	0	0	0
[6]	0	0	1	2	6	3	30	3	1
[7]	0	0	0	0	0	0	0	0	0
[8]	0	0	0	0	2	0	1	0	6

Key:

[0]	truth-func. denial
[1]	none
[2]	persp. dep. denial
[3]	negative agreement
[4]	rejection
[5]	negative imperative
[6]	mot. dep. denial
[7]	mot. dep. exclam.
[8]	self-prohibition

rel. obs. agreem.	0.563
Cohen's κ	0.461
$AC1$	0.514

Table 5.32: *Confusion matrix for robot negation types (left) and measures for intercoder reliability (bottom right)*

lower limit to draw tentative conclusions. On the scale of Rietveld and van Hout (1993) (quoted in Di Eugenio 2000), in the following referred to as *Rietveld's scale*, our κ value is taken to be on the lower end of moderate agreement ($.41 < \kappa < .6$). Within the psychiatric community a κ-value of 0.46 would be considered too low due to it being < 0.5 (Grove et al. 1981: quoted in Di Eugenio 2000). The relative agreement for all coded utterances amounts to 56.3%, whereas the relative agreement for all coded utterances containing pragmatic negatives amounts to only 43.24%, which indicates that latter utterances are a "source of trouble" in terms of intercoder agreement.

Intercoder Agreement: Felicity of Robot Negatives The intercoder agreement in terms of κ for the felicity or adequacy of the robot's negative utterances amounts to 0.41, which is the lowest κ-value of all judgments that the two coders made. The relative agreement on felicity of all negative utterances amounts to 65.33%, yet for utterances containing pragmatic negatives only it amounts to as few as 48.65%. This goes to show that also in terms of felicity utterances with pragmatic negatives were less agreed upon

compared to utterances containing lexical negatives. We will uncover some of the reasons for the overall high level of disagreement in terms of felicity in subsection 5.3.7.

		coder 2																	
		[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]
coder 1	[0]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	[1]	0	84	0	0	0	0	0	2	0	0	0	0	0	1	0	0	1	0
	[2]	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	[3]	0	0	0	17	0	0	1	1	2	0	0	0	0	0	0	0	0	0
	[4]	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
	[5]	0	0	6	0	0	36	0	0	0	0	4	0	0	0	0	0	0	21
	[6]	1	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
	[7]	0	0	0	0	0	0	0	10	4	1	8	0	0	0	0	0	0	1
	[8]	0	0	0	2	0	0	1	1	12	0	0	0	0	0	0	0	0	0
	[9]	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
	[10]	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	[11]	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
	[12]	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
	[13]	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
	[14]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
	[15]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	[16]	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
	[17]	0	0	0	0	0	4	0	2	0	0	0	0	0	0	0	0	0	53

Key:					
	[0] neg. persp. quest.	[6] neg. persp. assert.	[12] neg. of self-prohib.
	[1] truth-func. denial	[7] neg. agreement	[13] quoted negation
	[2] neg. tag question	[8] disallowance	[14] rej. of request
	[3] prohibition	[9] neg. imperative	[15] neg. promise
	[4] mot. dep. assert.	[10] neg. question	[16] truth-func. neg.
	[5] neg. mot. quest.	[11] ?	[17] neg. int. interpret.

rel. obs. agreem.	0.779
Cohen's κ	0.739
$AC1$	0.768

Table 5.33: *Confusion matrix for human negation types (top) and measures for intercoder reliability (bottom right)*

Intercoder Agreement: Participants' Negation Types In terms of the aforementioned scales, the κ-value of 0.74 for the coders' decisions regarding negation types of participants' utterances is sufficiently good. It is well within Krippendorff's boundaries,

within which tentative conclusions can be drawn, it is well above Rietveld's and van Hout's 0.61 mark, thereby indicating substantial agreement amongst the coders, and it is equally well above the 0.6 mark established within the psychiatric community.

Upon examining table 5.33 we can identify differing judgments as to whether particular negative utterances were *negative motivational questions* or *negative intent interpretations* as the most common source of disagreement between coders. These two types essentially differ from each other only in whether the producer of the utterance is deemed to expect an answer or not, i.e. if the utterance is taken to be an assertion or a proper question. Here the first coder and author has a strong tendency to classify as proper questions what the second coder considers to be assertions. As has already been mentioned in section 5.3.4, it can be very difficult to determine if an utterance that is produced with an interrogative prosodic contour or, more precisely, an utterance which a coder perceives as having an interrogative prosodic contour, is an actual question or not. This problem is especially prevalent in the case of one-word utterances such as "no" which are abundant within the given data. One-word utterances lack any grammatical features that could give a coder further cues as to whether they are proper questions or not.

One could therefore argue to fuse these two types into one. For this reason we calculated both κ as $AC1$ values for a new hypothetical taxonomy in which *negative motivational questions* and *negative intent interpretations* are fused into one type, with all other types being equal to the current taxonomy. Subsequently a modified confusion matrix was calculated based on the given one (cf. table 5.33) by simply position-wise adding the respective rows and columns, columns and rows 17 and 5 in table 5.33. In order for the resulting hypothetical confusion matrix to be considered valid or meaningful, we have to assume that the two coders would classify all those utterances, which they classified under the current scheme as either one of these two types, as being of this new hypothetical, fused type.

This assumption seems reasonable, but cannot be proven to hold, as a reduced taxonomy with modified definitions in the coding scheme could have unforeseeable side effects on how coders decide to classify any particular utterance. For example having simply one type less to consider, or a particular name choice for the new hybrid type, could all influence the decisions of the coders.

Nevertheless we undertook this synthetic exercise in order to estimate the numerical impact that such a fusion of types may have upon the intercoder agreement. The κ-value of the resulting confusion matrix amounts to 0.814, and the resulting $AC1$-value amounts to 0.849. Thus, if our assumption holds, we could expect a close to impeccable agreement between coders, as even the toughest of the abovementioned scales, Kripendorff's scale, sets 0.8 as the lower boundary for κ-values, starting from which definite conclusions can be drawn from the underlying classification.

Yet, due to the postulated significance of *intent interpretations* for human language acquisition, the abundance of this type in our experiments, and also due to the lack of man-power necessary to re-code the human utterance set with the modified taxonomy, we decided to keep our original taxonomy for participants' negation types. Nevertheless it is important to keep in mind, that negative variants of precisely these theoretically important *intent interpretations* can, according to our 2-coder analysis, be easily confused with *negative motivational questions* and vice versa.

The κ value of the 2-coder classification of the robot's negation types was considered too low and we therefore attempted to improve the taxonomy of the robot's negation types.

5.3.6 Automatic optimization attempt for the robot's negation taxonomy

Our first attempt to improve the taxonomy of the robot's negation types might be best described as an automatic optimization approach that seeks to reduce the number of types by merging one or more times two or more types into one resulting type. This approach is motivated by the hypothesis that the insufficient intercoder agreement described in section 5.3.5 might be caused by too fine distinctions between the various types. If this was the case, one would expect that a reduction of types by merging two or more types into one would result in an improvement of the intercoder agreement and therefore the intercoder reliability. In order to determine in a principled way those mergers which could be expected to lead to a sufficient or even excellent intercoder agreement an optimization program was implemented (cf. algorithm 3).

Taking our given confusion matrix as input the program performs an exhaustive search over all possible sets of mergers with κ as the to be optimized value. Furthermore we applied the constraint that the application of all mergers specified in any particular set yields a taxonomy with no less than 3 types.

Due to our existing confusion matrix serving as the only input for the program, this approach makes the following implicit assumption: Assume that our optimization program suggests to merge some type A and some other type B into a fused type AB. If then, with the original taxonomy, coder 1 decided that a particular utterance is of type A, and coder 2 decided that the same utterance is of type B, we assume that, given the modified taxonomy with the fused type AB, both coders would decide that the same utterance is of type AB. Again, this assumption seems to be straightforward but, as we already mentioned above, cannot be proven to hold due to possible side-effects that a modified taxonomy might bring along (cf. the short discussion of the last paragraph of the previous subsection 5.3.5).

Algorithm 3 Optimization algorithm for taxonomies based on Cohen's κ

```
1: c ← |C^org|
2: Calculate P such that ∀P_i ∀s_j ∈ P_i : |P_i| < c ∧ Σ_{j=1}^{m} |s_j| − 1 < (C − 3) ∧ |s_j| > 1
3: for all P_i ∈ P do {perform the mergers as specified by P_i}
4:    C ← C_org {Each set of mergers starts from the original confusion matrix}
5:    for all s_j ∈ P_i with s_j = {x_0, ..., x_k} do
6:       a ← x_0
7:       for all b = x_1, ..., x_k do
8:          C_{a:} = C_{a:} + C_{b:}
9:       end for
10:      for all b = x_1, ..., x_k do
11:         C_{:a} = C_{:a} + C_{:b}
12:      end for
13:      for all b = x_1, ..., x_k do
14:         remove C_{:b} from C
15:         remove C_{b:} from C
16:      end for
17:   end for
18:   d = c − Σ_{j=1}^{m} |s_j| − 1 {the remaining number of rows/types}
19:   Calculate κ for C
20:   H_d(C) ← κ
21: end for
22: for all H_d with i = 3, ..., c − 1 do
23:    Sort H_d in ascending order
24:    print H_d
25: end for
```

Notation Within algorithm 3 and the example below (cf. figure 5.6) the following notation is used. Let C^{org} denote the original confusion matrix which is given to the program as input, C some confusion matrix, $C_{i:}$ the ith row, and $C_{:i}$ the ith column of C. Let $|C|$ denote the number of rows or columns of a confusion matrix, i.e. for the confusion matrix in figure 5.32 $|C| = 9$. Furthermore let $P = \{P_i\}$ with $i = 1, \ldots, n$ be the set of all partitions, and $P_i = \{s_j\}$ with $j = 1, \ldots, m$ be a particular partition. Furthermore let $s_j = \{x_k\}$ with $k = 0, \ldots, l$, $x_k \in \mathbb{N}$, $l < |C|$ and $x_a \neq x_b \forall a, b$ be a set of indices. The indices refer to the dimensions of the confusion matrix. Let $|x|$ denote the size of either

P, any P_i, or any s_j, i.e. the number of partitions, index sets, or indices correspondingly. Finally a set of hash-maps $\mathcal{H}_i$ is used to store mappings between confusion matrices C with $|C| = i$ and their corresponding κ-coefficients.

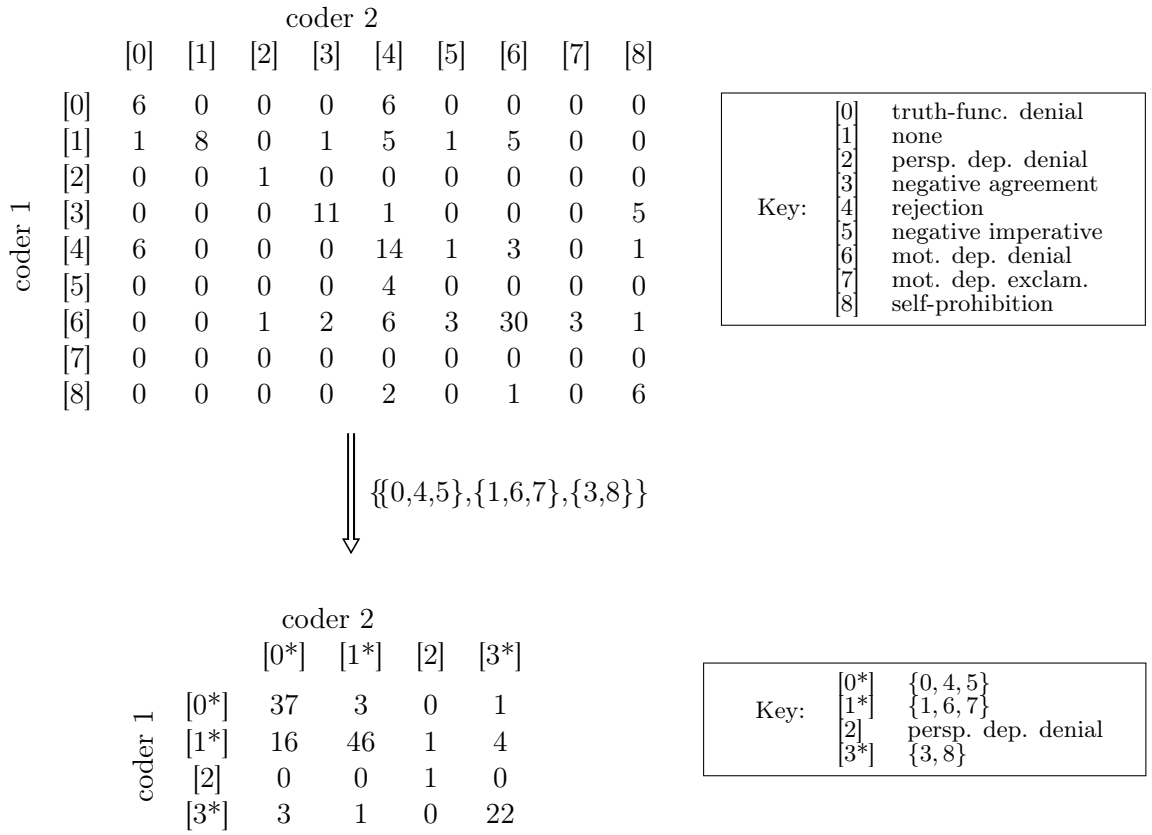


Figure 5.6: *Example of applying a set of mergers to the confusion matrix of robot negation types. Top left: The original confusion matrix C^{org} resulting from the two-coder coding of negative robot utterances. Bottom left: The resulting matrix when the various mergers indicated next to the arrow are applied to C^{org}. Dimension 0^* of the resulting matrix corresponds to the dimensions 0,4, and 5 of C^{org}. In other words, the potential new "negation type" resulting from the merger would be a hybrid type subsuming the "old" types* truth-functional denial*,* rejection*, and* negative imperative*.*

Example The partition $p = \{\{0, 4, 5\}, \{1, 6, 7\}, \{3, 8\}\}$ contains 3 sets and specifies therefore 3 mergers. The first set specifies the merger of the dimensions 0,4, and 5, i.e. the 4. and 5. row are added element-wise to the 1. row of the confusion matrix. Analogously the 4. and 5. column of said matrix are element-wise added to the 1. column. Subsequently the 4. and 5. row, and the 4. and 5. column are deleted from the confusion matrix such that the resulting matrix has 2 dimensions less in each rows and columns compared to the original matrix. The application of this set of mergers to the 9x9 confusion matrix obtained results in a matrix whose column and row vectors are reduced by 5 dimensions, i.e. a 4x4 matrix. This merger is visualized in figure 5.6.

Results from the optimization attempt

Tables 5.34 and 5.35 show those mergers with the highest resulting κ-values as calculated by our optimization program. The mergers are ordered by the highest resulting κ-values from top to bottom, and the number of types removed/merged from left to right. Table 5.34 shows the results obtained from the full negative utterance set, including pragmatic negation words (cf. subsection 5.3.3), table 5.35 shows the results from the negative utterance set, where pragmatic negation words were removed. Additionally both tables consist of two subtables each, one including and one excluding data from participant P12. The reason for P12's exclusion was his rather unnatural linguistic behaviour towards the robot, which was caused by his attempt to 'optimize the teaching game' by consciously reducing his vocabulary to the absolute minimum. P12 assumed that this would improve the robot's word learning. This information was relayed to us by a colleague, who is tenant and housemate of P12. This explained his rather 'minimalist' teaching style. Naturally such a conscious artificial adaptation of a participant's speech runs counter to our intention to evoke negation naturally by establishing a 'natural context' for such productions to occur.

We therefore decided to run our optimization program on the confusion matrices based on the data set containing P12's utterances and, additionally, on those confusion matrices based on the same data set but under exclusion of P12's utterances. This was done in order to be able to detect and potentially exclude any distortion of the intercoder agreement which could have possibly been brought about through P12's consciously unnatural speech style.

Indicated mergers for full data set (table 5.34, top subtable) With regards to their κ-value all of the top three mergers of the third column of the table are at the lower end of what is regarded as "substantial agreement" on Rietveld's scale. Interestingly the common denominator of all three mergers is the suggestion to fuse *truth-functional denial* with *rejection*, an operation which we are strongly opposed to due to reasons which will be discussed in subsection 5.3.8. Furthermore both the top- as the 3rd-ranking mergers suggest a further fusion of *truth-functional denials* and *rejection* with *negative imperatives*. The fusion of the latter two types is far less problematic as compared to the fusion of the first two types because *rejections* and *negative imperatives* are only distinguished from each other in terms of whether they are deemed to be in relation to a physical offer of the interlocutor or not, both are motivation-dependent types. Finally the top- and second-ranking mergers in the third column suggest a fusion of *negative agreement* with *self-prohibition*. Also this fusion seems from a theoretical standpoint not overly problematic, as the two types differ from each other mainly in terms of linguistic adjacency. If an utterance is judged to be a *negative agreement*, the judge sees it as an affirmative reaction to a previous negative utterance of the interlocutor. If an utterance is judged to be of the latter kind, the judge sees it as a stand-alone utterance in which the speaker seems to linguistically prevent him- or herself from doing something. In both cases the utterance is seen to express agreement

with a negative 'statement', often of the prohibitive kind, because of which both types are cognitively somewhat similar. A further interesting observation for this table is the circumstance, that virtually every merger from column two upwards suggests a fusion of *truth-functional denial* with *rejection*, and this same fusion is the one which yields the highest improvement in κ if only two types are merged overall (1st column, top row).

Indicated mergers for data set without P12 (table 5.34, bottom subtable) As was the case for the top subtable discussed in the previous paragraph, a reduction by at least 3 types is indicated by the optimization program, in order to yield a κ-value that would indicate substantial agreement on Rietveld's scale. By and large the indicated mergers are very similar to the ones that resulted from the optimization of the full data set discussed in the last paragraph: the most frequently indicated fusions within these mergers are the fusion of *truth-functional denial* and *rejection* on one hand, and the fusion of *negative agreement* and *self-prohibition* on the other. Yet, in contrast to the result from the top subtable, this optimization indicates an alternative merger, with which a fusion of *truth-functional denial* and *rejection* could be avoided: the top merger in the third column suggests that fusing *negative agreement* with *self-prohibition* on one hand and *rejections*, *negative imperatives*, and *motivation-dependent denials* on the other, would yield a sufficient κ-value. We discussed the former fusion already in the last paragraph. One could argue that the latter fusion is from a cognitive-developmental standpoint far less problematic than a fusion of *rejection* with *truth-functional denial* as these three types are very similar to each other. All three are linked to the motivational state of the agent that expresses them and do not seem to require the capacity for "logical abstraction" as seems to be the case for *truth-functional denial*. The difference between these three types is simply a matter of adjacency: *motivation-dependent denials* are adjacent to a linguistic

offer and *rejections* are adjacent to a non-linguistic offer, yet both are adjacent to some intentional move towards the speaker by another agent. *Negative imperatives* on the other hand are 'only' adjacent to an event and therefore do not require the presence of some intentional move. Yet utterances of all three types have in common that the speaker does not want something to happen, be it either brought about by somebody else or be it just an event that is not linked to any agent.

Indicated mergers for data set without pragmatic negatives (table 5.35, top subtable Notice that the taxonomy without pragmatic negatives has one type less, *none*, compared to the taxonomy based on the data set with pragmatic negatives because of which its dimensionality (number of types) is 8 and not 9. We kept the indices of the types within table 5.35 identical to the ones of table 5.34 in order to facilitate comparisons between tables. The reason for most mergers to contain one type less, is the circumstance that there is one type less in the taxonomy to start with.

Similar to the mergers indicated by both subtables in table 5.34 in the previous two paragraphs, the core types to be fused are *truth-functional denial* and *rejection* on one hand (all mergers of all columns from column 2 upwards), and *negative agreement* and *self-prohibition* on the other hand (from column 2 top merger upwards). As mentioned in the penultimate paragraph, a fusion of the latter two types into one hybrid seems not overly problematic, yet the merger of *truth-functional denial* with *rejection* seems very problematic and will be discussed in subsection 5.3.8. The indications which types ought to be fused such that substantial agreement between coders could be expected are basically identical to those regarding the two subtables discussed in the previous two paragraphs.

Indicated mergers for data set without pragmatic negatives and without P12's utterances (table 5.35, bottom subtable) The mergers for which there is an indica-

tion that substantial agreement between coders can be expected ($\kappa > 0.61$) start in column 2 with the top merger. Here, based on the smallest data set of all four, the optimization does not strongly indicate a fusion of *truth-functional denial* with *rejection*: only 3 out of 6 mergers (column 2 top merger plus all mergers from column 3) indicate such a fusion. Yet the fusion of *negative agreement* with *self-prohibition* is still indicated by 5 out of 6 mergers. Instead of fusing *truth-functional denial* with *rejection* the majority of mergers (5 out of 6) indicate a fusion of *rejection* with *motivation-dependent denial*. This fusion is from a cognitive standpoint one of the least problematic ones as the only difference between the two types concerns adjacency to either an utterance as opposed to a non-linguistic offer. As participants often physically offered an object to the robot whilst simultaneously asking if it wanted that object, it is easy to see how the coders could confuse these two types. One could even assert, that it might not be sensible to distinguish the two types from each other in our case precisely because physical offers were very often co-produced with linguistic ones, both of which fulfill arguably the same function[37].

[37] One could further argue that to ask a coder to which of these two intentional moves a robotic utterances is adjacent is to create an artificial problem which in the reality of the conversation does not occur: for the participant it does not matter if the robot rejects the physical or the linguistic part of his or her offer.

Table 5.34: *Optimization Results based on full intercoder data set and under exclusion of participant 12. Those mergers are marked bold from which on one could speak of sufficient intercoder agreement according to the scale of Rietveld and van Hout (1993)*

(a) *Optimization result on full intercoder data set*

Rank \ Dims. removed	1	2	3	4	5
(1) mergers	{0,4}	{0,4}{3,8}	**{0,4,5}{3,8}**	{0,1,4,5}{3,8}	{0,1,4,5,6}{3,8}
κ/AC_1	0.547/0.609	0.585/0.645	**0.623/0.680**	0.656/0.735	0.723/0.891
(2) mergers	{4,5}	{0,4,5}	{0,4}{3,8}{6,7}	{0,4,5}{3,8}{6,7}	{0,1,4,5}{3,8}{6,7}
κ/AC_1	0.497/0.549	0.583/0.645	0.610/0.661	0.650/0.693	0.686/0.746
(3) mergers	{3,8}	{0,4}{6,7}	{0,1,4,5}	{0,4,5}{1,6}{3,8}	{0,1,4,5,6,7}
κ/AC_1	0.495/0.549	0.572/0.623	0.610/0.701	0.642/0.717	0.685/0.874
(4) mergers	{6,7}	{0,4}{5,6}	{0,4,5}{6,7}	{0,1,4,5,6}	{0,4,5}{1,6,7}{3,8}
κ/AC_1	0.484/0.532	0.564/0.623	0.609/0.662	0.640/0.853	0.670/0.726
(5) mergers	{5,6}	{0,1,4}	{0,1,4}{3,8}	{0,1,4,5}{6,7}	{0,4,5,6,7}{3,8}
κ/AC_1	0.475/0.533	0.562/0.657	0.604/0.692	0.639/0.717	0.664/0.806

(b) *Optimization results based on intercoder data without P12*

Rank \ Dims. removed	1	2	3	4
(1) mergers	{4,6}	{3,8}{4,6}	**{3,8}{4,5,6}**	{0,4,5,6}{3,8}
κ/AC_1	0.535/0.656	0.590/0.709	**0.641/0.749**	0.694/0.787
(2) mergers	{0,4}	{3,8}{0,4}	{3,8}{0,4,6}	{0,1,4,6}{3,8}
κ/AC_1	0.535/0.591	0.582/0.642	0.640/0.749	0.682/0.832
(3) mergers	{3,8}	{4,5,6}	{0,4,5,6}	{1,4,5,6}{3,8}
κ/AC_1	0.531/0.605	0.580/0.697	0.627/0.736	0.681/0.832
(4) mergers	{5,6}	{0,4,6}	{0,4}{3,8}{5,6}	{0,1,4,5,6}
κ/AC_1	0.513/0.580	0.579/0.697	0.613/0.665	0.681/0.832
(5) mergers	{4,8}	{0,4}{5,6}	{0,1}{3,8}{4,6}	{0,1}{3,8}{4,5,6}
κ/AC_1	0.506/0.581	0.563/0.617	0.603/0.711	0.654/0.747

Key:	
[0]	truth-func. denial
[1]	none
[2]	persp. dep. denial
[3]	negative agreement
[4]	rejection
[5]	negative imperative
[6]	mot. dep. denial
[7]	mot. dep. exclamation
[8]	self-prohibition

Table 5.35: *Optimization Results based on intercoder data set without pragmatic negatives and under exclusion of participant 12. Those mergers are marked bold from which on one could speak of sufficient intercoder agreement according to the scale of Rietveld and van Hout (1993)*

(a) *Optimization results based on intercoder data without pragmatic negatives*

Rank \ Dims. removed	1	2	3	4
(1) mergers	{0,4}	**{0,4}{3,8}**	{0,4}{3,8}{6,7}	{0,4}{3,8}{5,6,7}
κ/AC_1	0.610/0.688	**0.672/0.740**	0.711/0.767	0.750/0.792
(2) mergers	{3,8}	{0,4}{6,7}	{0,4}{3,8}{5,6}	{0,4,5,6}{3,8}
κ/AC_1	0.546/0.616	0.646/0.716	0.709/0.768	0.726/0.884
(3) mergers	{6,7}	{0,4}{5,6}	{0,4}{5,6,7}	{0,4}{2,6,7}{3,8}
κ/AC_1	0.523/0.592	0.644/0.716	0.682/0.743	0.721/0.765
(4) mergers	{5,6}	{0,4}{2,6}	{0,4}{2,6}{3,8}	{0,4,5}{3,8}{6,7}
κ/AC_1	0.520/0.593	0.617/0.692	0.681/0.743	0.720/0.765
(5) mergers	{4,5}	{0,4,5}	{0,4,5}{3,8}	{0,4}{2,5,6}{3,8}
κ/AC_1	0.500/0.569	0.617/0.693	0.681/0.743	0.719/0.765

(b) *Optimization results based on intercoder data without pragmatic negatives and under exclusion of data from P12*

Rank \ Dims. removed	1	2	3
(1) mergers	{3,8}	**{3,8}{4,6}**	{3,8}{4,5,6}
κ/AC_1	0.549/0.655	**0.628/0.788**	0.722/0.848
(2) mergers	{0,4}	{3,8}{0,4}	{3,8}{0,4,6}
κ/AC_1	0.528/0.619	0.608/0.693	0.697/0.831
(3) mergers	{4,6}	{4,5,6}	{0,4,5,6}
κ/AC_1	0.524/0.712	0.602/0.772	0.667/0.814
(4) mergers	{5,6}	{3,8}{5,6}	{3,8}{0,4}{5,6}
κ/AC_1	0.513/0.621	0.594/0.695	0.657/0.730
(5) mergers	{4,8}	{0,4,6}	{3,8}{2,4,6}
κ/AC_1	0.499/0.623	0.582/0.755	0.637/0.797

Key:	
[0]	truth-func. denial
[1]	none
[2]	persp. dep. denial
[3]	negative agreement
[4]	rejection
[5]	negative imperative
[6]	mot. dep. denial
[7]	mot. dep. exclamation
[8]	self-prohibition

Summary of the results of the automatic optimization attempt We ran our automatic optimization program as described in algorithm 3 on four variants of the dual-coded dataset, which on its part consists of 20% of the utterances from the set of all negative utterances. The four variants of this dataset differ from each other in whether they include utterances that contain pragmatic negation words and in whether they include the utterances of participant P12. P12 received this special treatment because he displayed a very unnatural way of speaking and who is the only participant of whom we know that he consciously modified his way of speaking to the robot in order to 'win the game'. This conscious modification of his speech style ran counter to our instructions to speak to the robot as if it was a child of approximately 2 years. The outcome of running our optimization program on these datasets yielded the following:

For 3 out of 4 of these datasets, the complete dataset, the dataset with pragmatic negatives but without P12's utterances, and the dataset without pragmatic negatives but with the utterances of P12, the results were very similar: most of the mergers which would arguably result in a taxonomy with a substantial intercoder agreement contain a fusion of *truth-functional denial* with *rejection* and one or more additional types. Additionally most of these mergers suggest a fusion of *negative agreement* with *self-prohibition*. The latter fusion is from a theoretical cognitive standpoint little problematic, yet the former fusion would run counter to everything that the literature on the development of linguistic negation suggests. Out of the indicated mergers of these 3 datasets only the dataset with pragmatic negatives but without P12's utterances (table 5.34, bottom subtable) indicates that the fusion of *truth-functional denial* and *rejection* could be avoided while still maintaining the expectation of substantial intercoder agreement. This merger suggests a fusion of all but one motivation-dependent negation types, *rejection*, *negative imperative*, and *motivation-dependent denial*, into one hybrid type, on top of the already mentioned fusion

of *negative agreement* with *self-prohibition*.

The results of the optimization performed on the fourth and smallest dataset, i.e. the dataset which neither contains utterances with pragmatic negation words nor P12's utterances, finally supports the alternative fusion which avoids lumping together *truth-functional denial* and *rejection*. There, the suggestion is to fuse *rejection* with *motivation-dependent denial* in addition to the "standard fusion" of *negative agreement* with *self-prohibition* which is also indicated by most other mergers of all other datasets.

In summary we can say, that we could reasonably expect that two coders would reach substantial agreement if we modified our taxonomy in one of these ways, i.e. by conducting one of the discussed mergers. Before making suggestions for any concrete merger, we will discuss in the following subsection a more humanistic and non-automatic attempt to uncover the source of the observed low intercoder agreement for the negative utterances produced by the robot which complements this automatic approach.

5.3.7 Qualitative analysis of the taxonomy

Despite the current positioning of this subsection, i.e. after the subsection which covers the automatic optimization of the taxonomy, the procedure described here was performed before the program for the automatic optimization was written. The results of this procedure are therefore independent of the automatic optimization attempt and not influenced by its results.

In order to determine the reasons for the high level of disagreement between coders, the 2nd coder was interviewed in the following way.

First, we wrote a script to automatically print all file names and precise time stamps of those utterances where both coders disagreed either on the negation type or the felicity of the utterance. The script did not print the categories on which both coders decided,

such that the resulting list only indicated the utterances on whose category the coders disagreed but not the particular kind of disagreement. Based on this list the first coder thereafter wrote down his reasons for his particular decisions, separately for each indicated utterance. Additionally the first coder noted all competing types if there were any, as there were cases where he was torn when deciding for one of several types. Subsequently the author and first coder interviewed the second coder where, for each indicated utterance, he asked her for the reasons why she decided the way she did. Furthermore the second coder was also asked if she could imagine to categorize the utterance differently and if yes, which alternative category she would choose. She was also asked if the choice of an alternative category would have an impact on her judgment on the felicity of the utterance.

The interview provided us with the general insight, that a coder's judgments on the felicity of an utterance often depends on his or her judgment with regard to the utterance type, and that it did not matter whether the latter judgment is made explicit with the help of a coding table or if it remains implicit. Remember that we separated for the 2nd coder the coding for the felicity/adequacy of an utterance in stage 1 from the coding for the negation type in stage 2 (cf. subsection B.8.3 in the appendix). This separation was motivated by the ethnomethodologically motivated idea that a fluent English speaker should be able to decide if a particular utterance is felicitous or 'makes sense' in a particular situation without any explicit knowledge of formal criteria such as SAT-like satisfaction conditions but rather based on his or her capacity as being a competent member of the speech community. Due to the first coder in his function as designer of the taxonomy being acquainted with all negation types at the time of coding of both felicity and negation types, the partial dependency of felicity and negation type mainly transpired during the interview with the 2nd coder. There she clarified that she would have chosen for various utterances different felicity values if she would have been aware of the various negation types from

the very outset. In particular referring to *self-prohibition*, the 2nd coder made clear, that she had totally forgotten about this kind of behaviour engaged in by small children, and subsequently did not consider it when judging on the felicity of a particular utterance. Upon reading about *self-prohibition* in the coding scheme she then, in retrospect, would have often changed her decision on felicity in cases where *self-prohibition* was a likely candidate for an utterance.

Reasons for Disagreement amongst Coders

The given reasons were subsequently analysed and put into the following categories. The order in which these reasons are listed is roughly in terms of their generality, i.e. reasons that only apply to particular types are listed first, and reasons which potentially apply to a multitude of types are listed later. Coders 1 and 2 are in the following abbreviated with *C1* and *C2*, participants with *P*, and the robot Deechee with *D*.

[1] C2 interpreted *motivation-dependent denial* more narrowly as compared to C1: For C2 the 'triggering' utterance had to be a *motivation-dependent question* or assertion that directly referred to the motivational state of Deechee such as "Do you want X" or "Do you want to hold X". Assertions such as "I'm going to show you another box" or "Let's look at the box X" did not qualify for C2 as adequate first pair-parts for *motivation-dependent denials*, whereas they did for C1.

Affected types:

- *motivation-dependent denial*
- *negative imperative*

[2] C2 interpreted *negative imperatives* more generally as compared to C1 such that they included utterances that were deemed as being linguistically adjacent to an utterance

of P. Replies to assertions which, for C2, did not qualify as adequate first pair-parts under [1], were subsequently often classified as *negative imperatives*.

Affected types: see [1]

[3] C2 interpreted *self-prohibition* more generally as compared to C1 to include linguistically adjacent utterances.

Affected types:

- *negative agreement*
- *self-prohibition*

[4] C2 interpreted *rejection* more generally to include asocial triggers such as the mere presence of an object.

Affected types:

- *rejection*

[5] Categorical overlap between *motivation-dependent denial* and *negative agreement*.

Affected types:

- *motivation-dependent denial*
- *negative agreement*

[6] Negation within the naming game versus not wanting to play the game - the 'in game' *no* versus the 'meta' *no*. This includes cases where C2 decided during coding stage 1 that D is not playing the game properly when saying *no* instead of the object label expected by P. This then often resulted in C2 judging D's utterance as not felicitous. Yet, in coding stage 2, C2 thereupon often chose to categorize the utterance as *truth-functional denial*, which implies that D is playing the game as it would have to be categorized as *rejection* otherwise. Furthermore the coders sometimes decided differently in terms of the particular utterance of P to which D's utterance was judged

to be adjacent to. If D's utterance was deemed to be adjacent to an object label, the categorical outcome would typically be *truth-functional denial*, whereas if it was deemed to be adjacent to a motivation-dependent question, it was deemed to be a case of *rejection*.

Affected types:

- *rejection*
- *truth-functional denial*

[7] Disagreement with regards to P's perception of pragmatic negatives, including judgments whether a particular utterance of D was actually heard by P or not.

Affected types:

- *none*
- *negative agreement*
- *truth-functional denial*
- *motivation-dependent denial*
- *negative imperative*
- *rejection*

[8] Disagreement as to whether an utterance is linguistically adjacent to an utterance produced by P or not, other than [2] or [3]. This includes cases where both coders agree that D's utterance is linguistically adjacent to one of P's utterances but disagree about the particular utterance of P to which D's utterance is supposedly adjacent to.

Affected types:

- *negative agreement*
- *rejection*

- *motivation-dependent denial*
- *motivation-dependent exclamation*
- *negative imperative*
- *self-prohibition*
- *truth-functional denial*

[9] Disagreement as to whether D's utterance is adjacent to an offer by P or not. This reason for disagreement often had to do with the timing of D's utterance. If, for example, P offered an object to D, and what may be interpreted as D's linguistic response was produced several seconds later, C1 had a tendency to not judge it as a response to the offer as the time gap was deemed too long. C2 on the other hand often still judged D's production to be a response to P's offer.

[10] Disagreement with regards to the intention of D, especially if the body behaviour is not congruent with the linguistic behaviour as is the case when Deechee is smiling and saying "no", or when Deechee is smiling but not grasping due to a technical error. Here C1 due to being the programmer of the system, knew that the non-grasping was a technical error and assumed that Deechee 'wanted' the object. C2 on the other hand often took the non-grasping to mean, that Deechee despite its smiles really did not want a certain object. Furthermore differing judgments amongst the coders as to whether the dropping of a box by D constituted an intentional action or not fall into this category. This reason mainly applied to disagreements on felicity rather than to differing judgments on the negation type.

Frequencies of reasons for type disagreement Upon extraction of the above list we assigned to each case of disagreement one or more reasons of disagreement and counted them. For example reasons [8], disagreement in terms of linguistic adjacency, and [9],

disagreement in terms of 'behavioural' adjacency, often were judged to apply to the same utterance. One coder may decide that D's utterance was adjacent to a previous utterance of P, whereas the other coder may decide that D's utterance was adjacent to P's corporal offer, but not to what he or she said. Therefore, in terms of counting, the counter for each of the given reasons was incremented if a disagreement case had more than one reason associated with it. Table 5.36 depicts the results of this analysis.

Table 5.36: *Counts of reasons for disagreement with regards to negation types between coders:* ctc (cX)*: count type change coder X: number of type changes for reason [x] as indicated by coders*

reason	count	%	ctc (c1)	%ctc (c1)	ctc (c2)	%ctc (c2)	ctc (c1+2)	%ctc (c1+2)
[1]	2	2.9	0	0	0	0	0	0
[2]	5	7.1	1	4.3	0	0	1	2.6
[3]	5	7.1	0	0	4	25	4	10.3
[4]	1	1.4	0	0	0	0	0	0
[5]	2	2.9	1	4.3	0	0	1	2.6
[6]	10	14.3	4	17.3	6	37.5	10	25.6
[7]	13	18.6	2	8.7	1	6.3	3	7.7
[8]	24	34.3	13	56.5	4	25	17	43.6
[9]	6	8.6	1	4.3	1	6.3	2	5.1
[10]	2	2.9	1	4.3	0	0	1	2.6
total	70		23		16		39	

As can be seen from table 5.36 the majority of disagreements between coders are reason *[6]*, the "in game" *no* versus the "meta" *no*, reason *[7]*, disagreement as to whether a pragmatic negative is an actual negative in the particular situation, and reason *[8]*, disagreement with regards to adjacency, i.e. disagreement as to whether an utterance is linguistically adjacent or not, and if both deem this to be the case, possible disagreement with regards to the first pair-part. Reason *[8]* explains over a third of all observed disagreements.

Notice that disagreements caused by reasons *[1]* to *[5]* are comparatively easy to remedy by refining the definitions and descriptions of these types and possibly by training the 2nd coder on a small test set of utterances. Yet the most frequent reasons for disagreement, reasons *[6]* to *[8]* cannot be overcome by simply fusing, adding, or redefining types. Reason *[7]* is exclusively linked to utterances containing pragmatic negatives, which we added to the set of to be coded utterances only tentatively. Thus if we remove the latter from the utterance set as is also indicated by the automatic optimization in section 5.3.6, reason *[7]* will disappear. Yet reasons *[6]* and *[8]* will remain problematic factors for even a reduced utterance set. Both of them have to do with the coders having problems in deciding whether an utterance is adjacent to a previous utterance, and, if yes, to which one that would be. The obvious solution to these problems is an improvement of the robot's turn-taking skills. This is easily said but equates to a fundamental overhaul of at least the languaging system, as the robot has at least to know that it is being addressed and would have to react in no later than approximately one second.

Frequencies of reasons for disagreement on felicity Similar to the analysis of reasons for disagreements with regards to the pragmatic type, we tabulated the reasons for disagreements with regards to the felicity or adequacy of utterances in table 5.37. As can be seen there the most frequent reason for disagreements on felicity is reason [7], which just applies to utterances containing pragmatic negatives. As already mentioned in the last paragraph, this reason disappears as soon as we eliminate utterances containing pragmatic negatives from the coding set. Interestingly the second-most frequent reason is one of the most general reasons for disagreements between coders: disagreement with respect to the intention of Deechee, i.e. disagreement as to whether Deechee really wants an object or not or disagreement as to whether Deechee performed an action intentionally or not. This

reason did not play much of a role for coders when trying to determine the negation type, as can be seen in the last paragraph. Yet this result indicates that knowing what an agent is intending or not intending to do, may be very important in order to judge if an utterance is adequate in a particular situation, especially when it comes to negation. Technical glitches in the arm controller of the robot sometimes prevented it from reaching towards an object, when it actually wanted to, indicated by Deechee's smile. Coder 1, as developer of the system, was perfectly aware of this technical shortcoming, did not 'get fooled' by this and knew that Deechee, despite its physical handicap, still wanted to reach for the object and therefore judged its intention as such: Deechee wants the object. The 2nd coder though, who had not been involved with the technical development of the system, did not know about this and therefore at times judged Deechee's intention differently. This in itself is an interesting result: even when judging an utterance on felicity in an intuitive manner, i.e. without adhering to any formal criteria, utterances are still judged in the light of the assumed intention of the speaker. This is a prediction that speech act theorists would possibly have made, if they cared about any real-world data: the speaker attitudes and intentions are in SAT formal criteria that determine the felicity of an utterance. The remaining two most frequent reasons, [8] and [6], were also frequent reasons for disagreement on the type of a negative utterance and were briefly discussed in the previous paragraph.

Table 5.37: *Counts of reasons for disagreement with regards to felicity between coders*

reason	count
[1]	2
[2]	2
[3]	1
[4]	3
[5]	0
[6]	8
[7]	16
[8]	9
[9]	2
[10]	10

Table 5.38: *Potential changes of negation types as indicated by coders during review of utterances of disagreement. The order of the listing, type 1 ↔ type 2, does not the indicate the direction of correction, old type → new alternative type.* $[x] + [y]$ *denotes that two reasons for disagreement were given for a particular utterance of disagreement, a x [Y]: reason Y applied a times for this pair of types.*

coder 1			
type 1	type 2	freq.	reasons for uncertainty
mot-dep. denial	mot-dep. exclamation	4	3x[8],[8]+[9]
rejection	truth-func. denial	4	3x[6],[8]
rejection	mot. dep. denial	4	2x[8],[8]+[10],[2]
truth-func. denial	mot. dep. denial	2	[6][8]
none	neg. agreement	2	2x[7]
mot-dep. denial	neg. agreement	1	[5]
rejection	mot-dep. exclamation	1	[8]
self-prohibition	rejection	1	[8]
self-prohibition	mot-dep. denial	1	[8]
truth-func. denial	neg. agreement	1	[8]
coder 2			
type 1	type 2	freq.	reasons for uncertainty
rejection	truth-func. denial	6	6x[6]
self-prohibition	neg. agreement	5	4x[3],[8]
rejection	mot-dep. denial	2	[8],[9]
self-prohibition	mot-dep. denial	1	[8]
rejection	neg. agreement	1	[8]
mot-dep. denial	neg. imperative	1	[7]

Frequencies for alternative types for cases of disagreement as indicated by coders As briefly mentioned above, the coders also noted alternative type decisions upon reviewing their categorical judgments for utterances of disagreement. The outcome of this process is displayed in table 5.38, separately for each coder and with an additional listing of the reasons associated to each type-pair. Note that the order of the pairs in this table does not state, which of the respective two types in each row is the type which was effectively chosen and which one is the alternative that was indicated as such during the review.

Uncertainties of coder 1 Two of the three most frequent type-pairs for coder 1, together accounting for more than 50% of uncertainties on part of this coder, are uncertainties in terms of adjacency. The alternatives *mot.-dep. denial* ↔ *mot.-dep. exclamation* indicates that the coder was uncertain in these cases as to whether the robot's respective utterance was adjacent to an utterance by the participant (→ *mot.-dep. denial*), or not adjacent to any communicative move at all (→ *mot.-dep. exclamation*). The alternatives *rejection* ↔ *mot.-dep. denial* indicate as well that the coder was uncertain in terms of adjacency, yet in this case the coder decided that the robot's respective utterance is adjacent to some move, but uncertain as to whether it is adjacent to an utterance or to a non-linguistic offer. The third-most frequent type-pair, *rejection* ↔ *truth-functional denial* cannot be explained exclusively in terms of adjacency. This uncertainty appears to be of a more fundamental kind: the coder is not sure if the negative utterance is a move within the current 'naming game' such as "No, this is not a heart" (→ *truth-func. denial*) or if the speaker is refusing to play the naming game with this particular object. The latter, compared to the other utterance, is what one may call a meta-move in the sense that it has the potential to stop the naming game itself as opposed to being a move within the game. How caretakers manage to distinguish one from the other, and which communicative resources toddlers use to express each of these two fundamentally different moves, is to our knowledge not documented. We consider this a rather important issue. What the analysis here suggests is that the robot is lacking the important capacity to act communicatively such that the two communicative moves can be clearly distinguished from an external observer's viewpoint. We don't think that any change within the taxonomy could overcome this problem and would not be surprised to see similar problems amongst the participants, which might be uncovered by a full-fledged conversation analysis of the complete corpus.

Uncertainties of coder 2 One of the two most frequent type-pairs for coder 2 in terms of this coders uncertainty is *self-prohibition* ↔ *neg. agreement*. This uncertainty could be easily overcome by defining these two types more intelligibly and by a potential training of the coder. The most frequent type-pair in terms of uncertainty, *rejection* ↔ *truth-functional denial*, is the same one as has just been briefly discussed in the previous paragraph and hints towards a much deeper problem than could be solved by a mere 'cosmetic' change of the taxonomy.

Effect of type changes as indicated by the two coders It could be argued that the low agreement of the two coders as measured by κ may be lower than is actually the case due to the coders making over-proportionally unfavourable choices when choosing, more or less by chance, between one of typically two potential types as indicated by 5.38. In order to see what outcome could be expected if one assumes that the coders would have happened to choose the respective other types as indicated in the table, we performed this hypothetical change, and calculated the ensuing confusion matrix as well as the κ-value associated with this matrix. Table 5.39 depicts the outcome of this operation. As can be seen there the κ-value would have been considerably higher had the coders chosen the respective other type in those cases of disagreement in which they, within the review, indicated a competing type. Yet the ensuing κ-value is most probably positively skewed by virtue of the coders only having re-examined utterances where both coders had disagreed. If the whole utterance set would have been re-examined including those utterances where both coders agreed in their decisions, most probably some change from agreement to disagreement would have occurred. Based on this reasoning the κ-value in table 5.39 is most probably too high, yet it is still located at the lower end of acceptable agreement.

		coder 2								
		[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
coder 1	[0]	9 *+3*	0	0	0	1 *-5*	0	1 *+1*	0	0
	[1]	1	8	0	0 *-1*	5	2 *+1*	3 *-2*	0	0
	[2]	0	0	1	0	0	0	0	0	0
	[3]	0	0	0	19 *+8*	1	0	1 *+1*	0	0 *-5*
	[4]	3 *-3*	0	0	0	19 *+5*	0 *-1*	1 *-2*	0	0 *-1*
	[5]	0	0	0	0	4	0	0	0	0
	[6]	0	0	0 *-1*	1 *-1*	5 *-1*	4 *+1*	33 *+3*	0 *-3*	1
	[7]	0	0	1 *+1*	0	0	0	1 *+1*	3 *+3*	0
	[8]	0	0	0	0	0 *-2*	0	1	0	6

Key:	[0] truth-func. denial	[3] negative agreement	[6] mot. dep. denial
	[1] none	[4] rejection	[7] mot. dep. exclam.
	[2] persp. dep. denial	[5] negative imperative	[8] self-prohibition

r. o. a.	0.726
κ	0.663
$AC1$	0.695

Table 5.39: *Confusion matrix ensuing from hypothetical changes of type decision (top) and measures for intercoder reliability (bottom right). Top table: small numbers in italic refer to changes in the confusion matrix based on the hypothetical type decisions,* r.o.a.*: relative observed agreement*

5.3.8 Insights from Combining the Quantitative with the Qualitative Results

In order to better understand the sources of the relatively high level of intercoder disagreement when determining the negation type of a particular negative robot utterance (cf. subsection 5.3.5), we combine in this subsection the results of the quantitative approach described in subsection 5.3.6 with the results from the qualitative analysis from the previous subsection 5.3.7.

Whereas the quantitative results indicate which types ought to be merged such that we could expect a sufficient intercoder agreement, the qualitative results focus on the reasons why certain types were confused in the first place. In other words, the quantitative approach yields precise recommendations while being explanatorily blind, the qualitative approach gives us explanations without making precise suggestions as to which types to

merge. Notice that the suggestions of the quantitative approach, i.e. the particular types which ought to be merged, coincide with those types which show a high relative disagreement in the confusion matrix (table 5.32), i.e columns and rows with relatively high entries outside of the main diagonal of the matrix. Starting with the suggested main mergers from subsection 5.3.6, we will have a look at the reasons for the 'confusion' of or disagreement on the involved types amongst coders and discuss potential solutions.

Confusion of *negative agreement* with *self-prohibition* The majority of all mergers which can be expected to yield a sufficient intercoder agreement in both tables 5.34 and 5.35 suggest a merger of both these types. According to confusion matrix 5.32 the coders disagreed 5 times in total with regards to these two types. In these cases it was always the 2nd coder that chose *self-prohibition* when the 1st coder chose *negative agreement*, and it was always reason *[3]* that caused this disagreement. That means that the 2nd coder's interpretation of *self-prohibition* was that she judged robot utterances to be from this type even though they were adjacent to a participant's utterance. Thereby she effectively ignored that the type is listed as a non-adjacent type in the coding scheme. Subsequently one solution in order to make disagreement between coders in these cases less probable would be to train the coders on a test set of utterances to ensure that they internalised the taxonomy sufficiently. Another solution would be to merge these two types. Yet, negative agreements can be far more than only self-prohibitive adjacent utterances, i.e. utterances in which the robot/child uses a word when approaching a forbidden object which was previously used by the participant or caretaker in a prohibitive manner. Negative agreements can consist of any kind of negative word, independent of having been used previously in a prohibitive manner or not. Merging the two would therefore water down the distinction between two potentially important functions which are both part of Pea's

taxonomy.

Confusion of *rejection* with *motivation-dependent denial* (and *negative imperative*) This merger is suggested by the results depicted in table 5.34, bottom subtable, and table 5.35, bottom subtable, sometimes with and sometimes without *negative imperatives*. According to the confusion matrix in table 5.32, *rejection* and *motivation-dependent denial* were confused 9 times, *rejection* and *negative imperatives* were confused 5 times, and *motivation-dependent denial* and *negative imperatives* were confused 3 times. The reasons for the confusion of *rejection* with *motivation-dependent denial* was mainly reason *[8]*, i.e. the coders disagreed as to whether the robot's utterance was linguistically adjacent to a human utterance or not. Reason *[10]* was twice analysed as being partially responsible for the confusion, i.e. the coders judged the robot's intention differently. The reasons for the confusion of *rejection* and *negative imperatives* were mostly reasons *[8]* and *[9]*, i.e. the coders, again, disagreed as to whether the robot's utterance was linguistically adjacent to a human utterance or not, or, more generally, whether it was adjacent to a physical offer or not. Albeit a fusion of these two or three types seems from a cognitive perspective the least problematic of all suggested mergers[38] because they are mainly distinguished in terms of conversational adjacency, merging these types into one would cover up an important issue linked to the robot's capabilities. Due to the robot's insensitivity to the participants' speech and its timing in real time, it is not surprising that coders find it difficult to judge many of its utterances with regards to adjacency: technically, and from the robot's perspective, they are neither adjacent nor non-adjacent, because the robot is wholly unaware that it is spoken to.

[38]This is, if we take *cognitive* to denote mainly the mental capacities of a single agent in isolation while ignoring certain social skills which cannot be measured in isolating settings as is the case in intelligence tests and the like.

We do know that toddlers at the onset of speech are already somewhat skilled in terms of turn-taking, which means in particular that they do answer questions and react to requests with some regularity. Our robot on the other has no notion of turn-taking. Due to its real-time deafness, its inability to recognise a human's moving mouth, and its inability to recognize if it is being looked at or not, Deechee stands no chance of deliberately taking a turn at the right time. It is lacking virtually all high-level 'sensory channels' that could give it a clue whether its interlocutor is executing a conversationally relevant move or not.

Thus, instead of glossing over this important difference in the skill set of toddlers compared to our robot, we should take this uncertainty of the coder's qua external judges as an important hint that turn-taking or sensitivity to the communicative move of a conversation partner is a non-negligible skill which a conversational robot cannot do without. If 'it does without' some of the consequences are the ones we see here: conversation partners as well as external observers will have severe difficulties discerning particular communicative functions from each other. We have shown that if the timing of the robot utterances, relative to the humans' utterances, is outside of a certain acceptable variance, this will have a major impact on the intelligibility of the utterance. Thus, turn-taking skills are not just something that are 'nice to have', but they are indispensable for the acquisition of negation.

Confusion of *rejection* with *truth-functional denial* The confusion between these two types is probably the most worrisome of all observed confusions. Whereas it seems at least theoretically fairly obvious how to alleviate the confusions discussed in the previous two paragraphs, the confusion between *rejection* and *truth-functional denial* occurs to be more fundamental than the others. The capacity to engage in turn-taking properly, i.e. to know that a questions requires an answer and how to properly produce an answer in

terms of intonation, content etc., might help with some of the confusions of *rejection* and *truth-functional denial*. Two reasons for this confusion transpired during the analysis: Reason [6], the most frequent reason, pertains to the disagreement between coders as to whether a negative utterance was an 'in game' move, or whether is was a 'meta-move'. Less frequently reason [8], i.e. uncertainty about linguistic adjacency was associated with this type of confusion. This reason would hopefully disappear if we equipped the robot with the necessary conversational skills. Yet we are uncertain if this would suffice to make the uncertainty behind reason [6] disappear.

Developmentally the emergence of *rejection* and *truth-functional denial* in child language are typically some months apart. One reason for the frequent confusion of these two types in our experiments might be the somewhat unrealistic setup of our language game. An important difference between the situational contexts and (language) games on which Pea's data is based and the language game played between the robot and our participants is that the former are many and the latter is one or two. That is, Pea's mother-child dyads were observed and videotaped while "playing, feeding, bathing, and during other home activities" (Pea 1980). Our human-robot dyads were observed within precisely two settings which were, in principle, very similar to each other: both scenarios were word learning scenarios. In such scenarios *truth-functional denial* is extremely likely to occur by the very nature of the task at hand: naming may be conceived of as one example of saying what is and what is not the case, that is, stating the name of an object while the object is physically co-present and thereby expressing public knowledge on what this object is called. Naturally, one may deny that a proposed label is the correct label for the object at hand which amounts to *truth-functional denial*. On top of this rather truth-functional scenario, we overlaid the motivational level, that is, we equipped the robot with the capacity to first physically, and then, eventually, linguistically reject a particular object. It

is unclear if such a scenario is realistic in the sense that it may not occur very frequently in mother-child interaction.

The approximate equivalent of this scenario in a setting with mother-child dyads is most likely the following: The mother shows her son or daughter a picture book with animals or the like, points to the various things displayed in the book, names them, and says other things about them. Furthermore, she most probably will ask the child if it knows the names of these things, either mainly for educational purposes or just for the purpose of involving the child in the game. *Rejection* on the part of the child may not occur very often in this setting, while the naming game is being played, so to say. *Rejection* on the part of the toddler seems likely to either occur before the start of the game, and may prevent the game from being played in the first place, or it may end the game. Pea (1980) does not specify the situational context or the particular language game within which he observed the particular types of negation. Thus we have to speculate within which or during the onset of which particular language games *rejection* and *truth-functional denial* were produced jointly within a small time frame, if at all. The temporally tightly interleaved usage of *rejection* and *truth-functional denial* seems not very likely in the aforementioned picture-book setting to us. Yet this is precisely the setting into which we forced our participants and based on which the coders had to make their decisions. Where it seems likely that a parent would end the game, if faced with recurring opposition, our participants are somewhat forced to go on due to the simple fact that they were told that each session lasted 5 minutes and that they should also present objects to the robot which the latter does not like. The instruction was given in order to increase the probability of negative intent interpretations being produced by participants. As we will see in the next section, this led precisely to what we intended to happen, i.e. *negative intent interpretations* and *negative motivational questions* were produced by participants in abundance. Yet we

thereby evidently created a somewhat artificial scenario, which subsequently led to many of the robot's productions of negation to qualify as *truth-functional denials* due to our setting.

Nevertheless we strongly believe that children even in the one-word stage do have means, expressed via some form of rejection, to stop an ongoing intersubjective activity such as the naming game, and that these forms must be discernible from the 'in-game' form of negation simply because they would not be efficacious otherwise. Whatever behaviour of the child is involved in making these two types of negation discernible from each other to the interlocutor, it is these behavioural manifestations, linguistic or otherwise, that would enable the interlocutor as the observer to distinguish *rejection* from *truth-functional denial*. We do not think that merging these two types for the sake of a hypothetically sufficient intercoder agreement is a good idea. We take the apparent confusion of these two types in our experiments as an indication that the acquisition of negation on part of the robot is incomplete.

We discussed in the previous three paragraphs the issues that would arise out of each of the suggested changes in the taxonomy. The most frequently suggested change of merging *rejection* with *truth-functional denial* in particular would alienate our taxonomy from those taxonomies of negation that have been set up to depict the developmental trajectory of negation in early child language. We therefore decided that a change of the taxonomy to hypothetically increase the intercoder agreement would gloss over some fundamental shortcomings of the robot's conversational capabilities and would not be beneficial for our understanding of the topic. Subsequently, due to the low intercoder agreement, we will have to be careful with any counts that are based on the often confused types.

It should be noted at this point that intercoder agreements as low as ours which have

been incurred during the pragmatic coding of human child utterances, have been reported by some researchers. These will be discussed in chapter 6.

5.3.9 Pragmatic Analysis of Participants' Negation

In this subsection we will use the taxonomy for human negation presented in the previous subsection in order to determine which types of negation were produced within the rejection and prohibition experiments by our participants. We will then compare the two experiments with each other to detect differences in terms of these types. Subsequently, the pragmatic and word level will be linked in order to characterise the relationship between functional negation types and lexical negation words by looking at two different aspects. First, we will determine the relative production frequencies of the most frequent negation words within the most frequent negation types. Thereafter we will look at the saliency rates of these words when being produced within instances of said negation types. Lastly, after having determined said saliency rates, we will have a closer look at the two factors which impact on prosodic saliency. By comparing the utterance length of negative utterances of the most frequently occurring negation types with salient negative words on the one hand with the general utterance length of utterances of these types on the other, we will get a better understanding about the impact of this factor on saliency and thereby, indirectly, on the role of prosodic emphasis.

All counts are based on the codes of single utterances as judged by the first coder which contain one or more of the lexical negation words listed in table 5.1. As mentioned in the introduction to this chapter, one utterance can constitute more than one negation type, especially if the utterance consists of several clauses. In the majority of cases however one utterance corresponds to one negation type (cf. subsection 5.0.1 for examples).

Table 5.40: ***Frequency of participants' negation types - Rejection experiment****. Listed are the counts for all negation types of all participants and all sessions within the rejection experiment. The last column lists the total count for each type across all participants minus the counts of participant P04. This participant has to be factored out for the subsequent consideration of salient words as a different method for detecting salient words was used. '?' is not a negation type but indicates that the coder could not decide on a type for a given utterance due to the utterance being incomplete.*

	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	total	total w/o P04
neg. intent interpret.	2	23	36	49	49	18	25	1	19	9	231	208
neg. mot. question	0	24	30	11	72	43	20	9	8	4	221	197
truth-func. denial	18	35	2	1	3	0	9	0	45	39	152	148
neg. agreement	0	4	5	0	16	9	1	0	0	0	35	31
neg. tag question	0	2	5	7	8	0	7	0	2	0	31	29
neg. persp. assertion	0	4	0	1	1	8	6	0	3	3	26	22
mot. dep. assertion	0	3	0	0	5	0	12	0	4	0	24	21
truth-func. negation	0	1	0	0	0	2	4	0	7	0	14	13
neg. imperative	0	1	0	0	6	0	0	0	0	4	11	10
neg. question	0	3	3	0	1	0	0	0	0	0	7	4
apostr. negation	0	1	0	0	0	2	2	0	1	1	7	6
truth-func. question	0	0	0	0	0	0	0	0	0	2	2	2
neg. persp. question	0	1	0	0	0	0	0	0	1	0	2	1
?	0	0	0	0	0	1	0	0	1	0	2	2
rejection	0	0	0	0	0	0	0	0	1	0	1	1
mot. dep. exclamation	0	0	1	0	0	0	0	0	0	0	1	1
neg. promise	0	0	0	0	0	1	0	0	0	0	1	1
total	20	102	82	69	161	84	86	10	92	62	768	666

Rejection Experiment

Table 5.40 gives an overview of the frequencies of the negation types produced by the ten participants within the rejection experiment.

Most frequently produced negation types If all frequencies are accumulated across participants the most frequent types produced within the rejection experiment are *negative*

intent interpretation, *negative motivational questions*, and *truth-functional denial*, in this order. Moreover these three types are heading the table by a considerable lead: on average 82.63% of all negative utterances are of one of these three types, with 62.79% of participant P09's utterances at the lower end and even 100% of participants' P01 and P10's utterances being of these types at the upper end of the distribution. If we abstract from the accumulated numbers, which represent the more talkative participants over-proportionally, rank the types according to frequency, and just consider the ranks, the results are not fundamentally different: *Truth-functional denial* ranks first amongst 4 participants, *negative motivational questions* and *negative intent interpretations* each rank first amongst 3 participants. With regards to the second-highest ranking types, *negative intent interpretations* occupy this place amongst 6 participants, and *negative motivational questions* is the second-highest ranking type amongst 4 participants. On the third rank the number of types increases to include *negative agreement*, on this rank for 3 participants, *negative tag questions*, on the 3rd rank amongst 2 participants, as are *negative motivational questions*. Also *negative imperatives* and *motivation-dependent assertions* occupy the third rank of one participant's frequency list each together with the already high-ranking *negative intent interpretations*.

Salience rates of the most frequent negation types Table 5.41 displays the salience rates of negation words for each utterance type, i.e. the percentage of negative utterances associated with each type whose negative words are the prosodically most salient ones. It is important to remember that our word extraction algorithm described in section 3.6 only extracts the prosodically most salient word from each utterance, and that these extracted words in their entirety subsequently form the basis of the robot's active vocabulary. In order for the robot to acquire any negative words, it is not sufficient that participants

produce utterances that contain negative words. The second requirement for negative words to become part of the robot's vocabulary is for them to be prosodically salient within the participants' speech.

As can be seen in table 5.41, the average salience rates of negative words within both *negative intent interpretations* and *negative motivational questions* occur to be higher than the same rate within *truth-functional denials*. The total rate across participants displayed in the table is calculated such that each utterance is given the same weight as opposed to giving each participant the same weight. This naturally skews the total towards the salience rates of participants who produced more utterances of the respective type. Adjusting the calculation such that each participant is given the same weight when calculating the total does not change this observation: The total salience rates resulting from this calculation for *negative motivational questions*, *negative intent interpretations*, and *truth-functional denial* are 40.32%, 33.01%, and 28.88% respectively. If, additionally, those salience rates based on very small frequencies are excluded, in our case frequencies smaller than 5, and each participant is given the same weight, the resulting salience rates for *negative motivational questions*, *negative intent interpretations*, and *truth-functional denials* are 44.33%, 36.2%, and 33.0% respectively. If we further exclude P04 from the calculation, a participant where a different word extraction method was used (cf. section 3.6), the resulting average salience rates for the three types in the same order as previously are 43.51%, 46.54%, and 32.86% . Thus all indications are that both motivational negation types have more salient negative words as compared to *truth-functional denial*. Yet, as some of the percentages are based on utterance sets of sizes < 10 we refrain from any definite assertion in this regard.

Truth-functional denial The type of negation which we may naturally expect to occur in a teaching scenario is *truth-functional denial*: it seems likely that utterances of

this kind would be produced whenever the pupil, in our case a robot, produces the wrong answer. It seems, so to say, part of the teaching game to correct the pupil when he gets it wrong. If, for example, the robot says "star" upon having been asked what the presented object is being called, and when the presented object is indeed a triangle, it would not surprise us to hear the participant qua teacher saying "No, it's not a star" or "No, it's a triangle", both cases of *truth-functional denial*. And indeed, as we can see from table 5.40, this kind of of utterance is produced many times by our participants within the rejection experiment, up to 72 times by participant P07 during the 5 times 5 minutes of all sessions. Only one single participant, P10, never produced *truth-functional denial*, the 9 other participants did. What is rather surprising in this context is that this utterance type was not produced more often by the participants of Saunders' experiment (cf. section 5.1.6) as both experiments were, as far as participants knew, about teaching the robot object names, and additionally, in Saunders' experiment, object colours and sizes.

Negative intent interpretations* and *Negative motivational questions As we know from the evaluation of the taxonomy (cf. section 5.3.5) *negative intent interpretations* and *negative motivational questions* were frequently confused amongst coders and are distinguished from each other only in terms of the former being judged to be an assertion and the latter being judged to be a question. We observed frequently that participants produced what the first coder judged to be *negative motivational questions* despite them clearly seeing the emotional display of Deechee. Furthermore they frequently did not leave the obligatory pause after asking these seeming questions, such that if was unclear to the coder if these utterances, which had the intonational contour typically associated with questions, were indeed that: proper questions. In such cases the two types are very easily confused and it is questionable if it makes sense in these situations to distinguish asser-

tions from questions. More important than the question of which of these two types these

Table 5.41: ***Percentage of negation types with salient negative word - Rejection experiment****. Listed are the percentages of utterances, classified by coder 1 as the stated negation type, (one of) whose negation words were detected as being salient relative to the total number of utterances of this type. All numbers are percentages relative to the total counts given in table 5.40. The last column lists the average percentage of salient negation words across participants minus participant P04. For participant P04, one of the first participants, a different algorithm for detecting salient words has been used. '?' is not a negation type but indicates that the coder could not decide on a type for a given utterance due to the utterance being incomplete. The* total *was calculated by weighing each utterance identically which effectively gives more weight to the salience rates of speakers who produced more utterances of the respective type.*

	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	total	total w/o P04
neg. mot. question	0	8.3	56.7	63.6	70.8	48.8	30	22.2	12.5	50	49.3	54.3
neg. intent interpret.	0	4.3	69.4	51	42.9	44.4	48	0	36.8	33.3	44.2	48.6
truth-func. denial	44.4	0	0	0	66.7	0	55.6	0	48.9	15.4	28.3	29.1
neg. agreement	0	0	100	0	87.5	44.4	100	0	0	0	68.6	77.4
neg. tag question	0	0	80	28.6	50	0	42.9	0	100	0	48.4	51.7
neg. persp. assertion	0	0	0	100	100	50	33.3	0	100	33.3	46.2	54.5
neg. question	0	0	100	0	100	0	0	0	0	0	57.1	100
neg. imperative	0	0	0	0	0	0	0	0	0	75	27.3	30
mot. dep. assertion	0	0	0	0	0	0	16.7	0	0	0	8.3	9.5
truth-func. question	0	0	0	0	0	0	0	0	0	100	100	100
rejection	0	0	0	0	0	0	0	0	100	0	100	100
neg. persp. question	0	100	0	0	0	0	0	0	0	0	50	0
?	0	0	0	0	0	100	0	0	0	0	50	50
apostr. negation	0	0	0	0	0	0	0	0	0	100	14.3	16.7
truth-func. negation	0	0	0	0	0	50	0	0	0	0	7.1	7.7
mot. dep. exclamation	0	0	0	0	0	0	0	0	0	0	0	0
neg. promise	0	0	0	0	0	0	0	0	0	0	0	0
total	40	3.9	65.9	50.7	58.4	46.4	36	20	39.1	29	41.8	47.6

utterances are instances of is that virtually all participants of the rejection experiment felt compelled at least once to refer linguistically to the negative motivational state of the robot

by producing one of these types without having been asked to do so[39]. Even more remarkably, 7 out of 10 participants, P04 to P10, produced these two types of negative utterances considerably more frequently than they produced *truth-functional denials* such that these negation types come to constitute the major source of negative words in the Rejection Corpus (cf. section 5.2.1). When designing the robot's behaviour, especially when designing its motivational displays, we were hoping that these behaviours would elicit these kind of utterances. Nonetheless we were surprised by the sheer abundance with which they were produced by participants without them having been asked to do so. Furthermore there are no indications that these utterances, at least on the surface, were linked to the teaching task that participants were given.

Prohibition Experiment

Table 5.42 displays the frequencies of the various negation types produced by the participants of the prohibition experiment.

Most frequently produced negation types Within the prohibition experiment a new negation type enters the top-ranking types, *prohibition*, thus leaving us with four top-ranking types. The fact that this type is leading the total frequency count is even more outstanding considering that the prohibition task was only in place during the first 3 sessions, i.e. nearly all of these linguistic *prohibitions* were produced during the first three sessions. As opposed to the rejection experiment the frequency drop between the top-ranking and the remaining types, i.e. the drop between the 4th- and the 5th-ranking type is not as pronounced: here the frequency drops by only 35.6%, whereas within the

[39]This is not to say that participants did not refer to its positive motivational state, indeed they did. Yet we only analysed negative utterances and these typically refer to the negative motivational state of the robot. We did observe many positive intent interpretations but can't give a numerical frequency due to our focus on negation within this thesis.

rejection experiment the observed drop is 79% between the 3rd and 4th-ranking type. The circumstance that the type frequencies are slightly more evenly distributed within the prohibition experiment can also be seen when looking at the percentages of utterances judged to be of these top-ranking types in relation to the total number of negative utterances: across participants on average 72.78% of all negative utterances fall into one of these four categories. If we add the frequencies associated with *disallowance* to this count, which is a type very similar to prohibition, the total coverage of these five types rises to 78.59% which is still lower as compared to the 82.63% covered by *negative motivational questions*, *negative intent interpretations*, and *truth-functional denials* of the rejection experiment.

Similar to the rejection experiment not much changes if we just consider the frequency ranks for each participant as opposed to the absolute frequencies. This ensures that the frequencies associated with the more talkative participants don't overly skew the totals. For three participants *prohibition* is the highest-ranking type, for two participants each *negative intent interpretations*, *negative motivational questions*, and *truth-functional denials* are the highest-ranking types, and *disallowance* is the top-ranking type for one participant. When looking at the second-ranking types, *negative intent interpretations* lead the table with being on this rank for 3 participants, followed by *prohibition*, *truth-functional denial*, and *negative tag questions* ranking 2nd for 2 participants each. The distribution of the third ranks is led by *prohibition* and *negative intent interpretations* with three participants each, followed by *negative tag questions* with two participants, and *negative agreement* with one participant.

We will thus focus in the following on the four top-ranking types plus *disallowance* as they cover more than three quarters of all negative utterances produced by participants. For the discussion of *negative intent interpretations* and *negative motivational questions* we refer to their discussion in the previous subsection on the rejection experiment. *Prohibition*

and *disallowance* will be discussed in the paragraph after next.

Table 5.42: ***Frequency of participants' negation types - Prohibition experiment.*** *Listed are the counts for all negation types of all participants and all sessions within the prohibition experiment. '?' is not a negation type but indicates that the coder could not decide on a type for a given utterance due to the utterance being incomplete.*

	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	total
prohibition	22	16	39	18	31	13	14	24	16	7	200
neg. intent interpret.	15	0	38	31	13	30	18	22	4	2	173
neg. mot. question	12	0	52	15	7	14	12	38	20	0	170
truth-func. denial	21	3	3	0	6	2	22	4	7	36	104
neg. tag question	1	0	14	30	0	16	1	2	0	3	67
disallowance	0	0	14	4	0	2	3	15	26	1	65
truth-func. negation	0	0	9	12	0	18	4	6	1	0	50
neg. agreement	0	0	15	3	10	0	7	3	5	0	43
mot. dep. assertion	0	0	3	12	1	1	0	5	1	0	23
neg. persp. assertion	1	1	0	4	0	5	1	2	0	1	15
negating self-prohibition	0	0	0	1	0	0	1	1	4	0	7
apostr. negation	0	0	1	1	0	1	0	0	3	0	6
neg. imperative	0	0	0	4	0	0	0	0	1	0	5
rejection	0	0	0	1	0	0	0	0	3	0	4
neg. question	0	0	0	2	2	0	0	0	0	0	4
neg. promise	0	0	2	0	0	0	0	1	1	0	4
neg. persp. question	1	1	0	0	0	0	1	0	0	0	3
?	0	1	1	0	0	0	0	0	0	0	2
mot. dep. exclamation	0	0	0	1	0	0	0	0	0	0	1
total	73	22	191	139	70	102	84	123	92	50	946

Salience rates of the most frequent negation types Table 5.43 displays the global salience rates of the various negation types produced by participants within the prohibition experiment. Similarly to the rejection experiment the total salience rates seem to indicate that the motivational types, if we count *prohibition* as a motivational type, have higher salience rates as compared to *truth-functional denial*. Also here the total for each type is

calculated based on each utterance being given the same weight such that the total represents the salience rates of more talkative participants more than the same rates of more taciturn participants. If we adjust for this, by just averaging across the salience rates of all participants, independent of their number of utterances of each type produced, the indicated differences are less pronounced. If we consider the salience rates of all participants, independent of the size of the basis, the total salience rates for *prohibition*, *negative motivational questions*, *negative intent interpretations*, and *truth-functional denial* are 58.96%, 42.74%, 45.26%, and 19.93% respectively. If we perform the same calculation but restrict it such that only the salience rates of those participants with a sufficiently large basis are considered, in our case participants whose number of utterances associated with the respective type is > 5, the salience rates of *prohibition*, *negative motivational questions*, *negative intent interpretations*, and *truth-functional denial* amount to 58.96%, 36.6%, 36.8%, and 29.2% respectively.

Thus the difference in salience between *negative motivational questions* and *negative intent interpretations* on one hand and *truth-functional denial* on the other hand is, according to the last method of calculation not as stark as the totals presented in table 5.43 seem to indicate, yet there still is a pronounced difference. The numerical basis for the latter calculation is not large enough though in order to draw any definite conclusions.

However, no matter which method of calculation we choose, within *prohibition* negation words reach the highest salience rate amongst all negation types. More than every second negation word is prosodically salient there.

Prohibition* and *Disallowance Linguistic *prohibition* and *disallowance* might be some of the least surprising negation types to be observed within this experiment. Within the prohibition scenario participants were told to prevent the robot from touching certain

Table 5.43: ***Percentage of negation types with salient negative word - Prohibition experiment.*** *Listed are the percentages of utterances, classified by coder 1 as the stated negation type, (one of) whose negation words was detected as being salient. All numbers are percentages relative to the total counts given in table 5.42. '?' is not a negation type but indicates that the coder could not decide on a type for a given utterance due to the utterance being incomplete.*

	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	total
prohibition	50	87.5	43.6	88.9	74.2	38.5	57.1	54.2	81.3	14.3	60.5
neg. mot. question	33.3	0	48.1	20	85.7	28.6	41.7	39.5	45	0	41.8
neg. intent interpret.	20	0	44.7	32.3	46.2	33.3	44.4	36.4	50	100	38.2
truth-func. denial	19	33.3	0	0	16.7	0	31.8	0	28.6	50	31.7
neg. tag question	0	0	35.7	53.3	0	56.3	100	50	0	33.3	49.3
disallowance	0	0	28.6	75	0	0	100	13.3	57.7	0	41.5
neg. agreement	0	0	46.7	33.3	60	0	71.4	66.7	80	0	58.1
truth-func. negation	0	0	11.1	41.7	0	27.8	25	0	0	0	24
mot. dep. assertion	0	0	33.3	41.7	0	0	0	0	0	0	26.1
neg. persp. assertion	0	100	0	8.3	0	0	100	50	0	0	17.4
rejection	0	0	0	50	0	0	0	0	66.7	0	75
neg. question	0	0	0	100	50	0	0	0	0	0	75
negating self-prohibition	0	0	0	0	0	0	0	0	50	0	28.6
neg. imperative	0	0	0	50	0	0	0	0	0	0	40
apostr. negation	0	0	0	0	0	100	0	0	33.3	0	33.3
neg. persp. question	0	100	0	0	0	0	0	0	0	0	33.3
?	0	100	0	0	0	0	0	0	0	0	50
neg. promise	0	0	0	0	0	0	0	100	0	0	25
mot. dep. exclamation	0	0	0	0	0	0	0	0	0	0	0
total	30.1	81.8	40.3	46.8	61.4	33.3	46.4	35	54.3	44	43.7

forbidden objects and they were taught how to achieve this in a physical manner, i.e. by pushing its arm away. Participants were not explicitly told to use any form of linguistic prohibition to achieve this feat, but intuitively we expected them to do precisely this. It seems rather typical for humans from European cultures to accompany their physical deeds, especially in the case of joint tasks, with fitting linguistic deeds, be it descriptions, prescriptions, commands, questions, and the like. Furthermore it is hard to imagine, at least within our European cultural circle, how one would go about prohibiting somebody

else by the use of words without using any negatives. We can imagine stopping somebody else from doing something, at least for a few seconds, by inarticulate, loud yelling. And if this happens contingently when he or she engages in some action, he or she would at some point understand that we try to do precisely this: stop them from doing something. Yet there seem to be other cultural norms at play, that typically prevent us from doing this, and language in such a situation is the typical tool of choice. For these reasons we did expect participants to engage in linguistic prohibition without us experimenters having told them to do so.

In English one would typically expect utterances such as "No!", "Don't!", "Don't do that!", "No, don't!", "No, not that!" or the like. We could also imagine a simple "Leave it alone!", or "Stop!", which may or may not be 'prefixed' by a "No". As we did not manually analyse the non-negative utterances of the prohibition corpus we cannot say precisely how many of these positively formulated prohibitions were uttered by participants, but we are certain that there are very few of them. As we repeatedly looked through the list of all words for lexical *and* pragmatic negatives before starting the pragmatic analysis we would be rather surprised if we would have missed these expressions. We are certain that there is no "leave it alone" in the corpus, and that there are only two "stop saying one"'s, that are not prefixed by a "no". Yet these were not produced in the context of our specified prohibition task and, therefore, do not fall into our *prohibition* category. Remember that this category was specifically set up to capture those forms of linguistic prohibitions which were used in the context of this task (cf. section B.8.5). Moreover, the above mentioned "stop saying X" was only produced two times by a single participant and is therefore exceptional.

Based on our corpus we can therefore state with sufficient certainty that linguistic *prohibition* and *disallowance* have very high negation densities as they were produced by

our participants, i.e. the probability that a prohibitive utterance or an utterance that disallows the robot something contains a lexical negative is extremely high. With regard to prohibitive utterances and disallowances that are performed in the context of our prohibition task indeed every single instance contained at least one negative word. This is considerably higher than the probability of an intent interpretation being negative or a motivational question being negative, where we have seen many positively formulated variants or where the lexical negative is replaced by what we called pragmatic negatives. This, combined with the circumstance that negative words have a comparatively high salience rate within *prohibitions* means that *prohibitions* are principally an excellent source of negative words for a word learner.

Comparison: Rejection vs. Prohibition Experiment

In order to attempt to answer our research question as to where the earliest forms of negation ultimately derive from, we compare in this section the two experiments with regards to differences in the number of productions of the various negation types produced by our participants. For this purpose we will analyse the first three sessions and the last two sessions as separate blocks as the rejection and prohibition experiments only differ from each other in the first three sessions (cf. chapter 4). We will focus on utterances which were categorised as instances of one of the four most frequent negation types, *negative intent interpretations*, *negative motivational questions*, *truth-functional denials*, and *prohibitions*, plus *disallowances*. The remaining, less frequent negation types were produced too infrequently in order to warrant a statistical analysis.

Table 5.44 lists the accumulated counts for the various types. The most conspicuous difference between the two experiments is the presence of linguistic *prohibition* and *disallowance* in the prohibition scenario and the absence of utterances of these types in the

Table 5.44: ***Counts of 5 most frequent negation types grouped by session blocks.*** *Listed are the counts of negative utterances belonging to one of the 5 most frequent negation types and grouped by sessions. As the rejection and prohibition experiments only differ during the first 3 sessions the type counts are summed across the first 3 session on one hand and across the last 2 session on the other for each respective experiment.*

	Rejection Experiment										
	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	total
					Sessions 1-3						
neg. intent interpret.	2	17	30	25	35	10	18	1	19	9	166
neg. mot. question	0	15	19	6	38	27	16	3	6	4	134
truth-func. denial	7	14	0	1	2	0	0	0	28	16	68
					Sessions 4+5						
neg. intent interpret.	0	8	6	24	14	8	7	0	0	0	67
neg. mot. question	0	9	11	4	34	16	4	6	2	0	86
truth-func. denial	11	22	2	0	1	0	9	0	17	23	85
	Prohibition Experiment										
	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	total
					Sessions 1-3						
prohibition	22	16	39	18	31	13	14	24	16	7	200
neg. intent interpret.	12	0	23	13	9	25	13	14	1	2	112
neg. mot. question	6	0	25	7	3	12	7	19	9	0	88
truth-func. denial	20	2	0	0	5	1	21	3	3	26	81
disallowance	0	0	14	4	0	2	3	15	26	1	65
					Sessions 4+5						
neg. intent interpret.	3	0	15	18	4	5	5	8	3	0	61
neg. mot. question	6	0	27	8	4	2	5	19	11	0	82
truth-func. denial	1	1	3	0	1	1	1	1	4	10	23

rejection scenario. This is not surprising as argued in the previous subsection as non-linguistic prohibition was part of the task in the prohibition scenario whereas it was not part of the participants' task in the rejection scenario.

When comparing the utterance counts of the negation types common to both experiments, the differences between the two experiments are far less conspicuous. A manual comparison of table 5.44, especially when comparing the totals, seems to indicate that

both *negative intent interpretations* and *negative motivational questions* were produced more frequently within the first three sessions of the rejection experiment as compared to the same sessions of the prohibition experiment. It could be argued that because participants engaged in linguistic *prohibition* and *disallowance* within the prohibition scenario they simply had less opportunity to engage in the other motivation-related negation types.

Table 5.45 shows the results of the statistical analysis, a t-test based on the counts displayed in table 5.44 comparing the respective types in both experiments against each other. As the length of the corresponding sessions in the prohibition experiment are slightly longer than the respective sessions in the rejection experiment, we performed the t-test on the unmodified counts (2nd column) as well as on adjusted counts (3rd column). In the latter case the corresponding counts from table 5.44 were divided by the factors by which the respective sessions in the prohibition experiment were longer than in the rejection experiment. The total duration of the sessions 1 to 3 of all participants within the rejection experiment is 8600.1 seconds, and the total duration of the same sessions of all participants within the prohibition experiment is 9361.3 seconds. Thus the adjustment factor by which the respective counts of the prohibition experiment are divided is $9361.3/8600.1 = 1.089$. The adjustment factor for the sessions 4 and 5 is calculated accordingly with the total length of all sessions in the rejection experiment being 5882.1 and the same length of the prohibition experiment being 6252.6. As can be seen from the table there is a strong tendency for lower means of production frequencies of ($p < 0.1$) for *negative intent interpretations* in the prohibition experiment as compared to the rejection experiment if the adjusted counts are considered as basis for comparison. In the case of *negative motivational questions* and of the unadjusted counts of *negative intent interpretations* the difference in means is also not statistically significant but there are still indications for this being the case ($p < 0.15$ and $p < 0.2$ respectively).

Table 5.45: ***Statistical comparison of most frequent negation types.*** *Displayed are the mean, standard deviation (sd) and t-values for the production frequencies of the displayed negation types based on the counts in table 5.44. The third column displays the respective means for the adjusted counts of the prohibition experiment. The adjustment factors are 1/1.089 for the first to third session and 1/1.063 for the fourth and fifth session.* $\star = p < 0.05$, $* = p < 0.1$, $\dagger = p < 0.15$, $\ddagger = p < 0.20$

	Rej. Exp.	**Pro. Exp.**		**Pro. Exp. (adj)**	
negation type	*mean (sd)*	*mean (sd)*	*T*	*mean (sd)*	*T*
		Sessions 1 - 3			
negative intent interpretation	16.6 (11.33)	11.2 (8.59)	$1.201^{\dagger}$	10.28 (7.89)	1.447^{*}
negative motivational question	13.4 (12.09)	8.8 (8.02)	$1.002^{\ddagger}$	8.1 (10.05)	$1.188^{\dagger}$
truth-functional denial	6.8 (9.59)	8.1 (10.05)	0.296	7.44 (9.23)	0.152
		Sessions 4 + 5			
negative intent interpretation	6.7 (7.72)	6.1 (6.01)	0.194	5.74 (5.65)	0.318
negative motivational question	8.6 (10.25)	8.2 (8.72)	0.094	7.71 (8.2)	0.213
truth-functional denial	8.5 (9.35)	2.3 (2.95)	$2.0^{\star}$	2.16 (2.77)	$2.055^{\star}$

Furthermore there is a statistically significant difference in the means of production rate of *truth-functional denials* in sessions 4 and 5 between the two experiments: participants of the rejection experiment produce significantly more *truth-functional denials* (8.5) as compared to participants of the prohibition experiment (2.3). In contrast there is no such difference between the two experiments with regards to the motivation-dependent types.

We can only speculate why participants engaged in so many fewer *truth-functional denials* in the prohibition experiment. It is possible that the presence of the prohibitive task with the first 3 sessions pushed the word-learning task somewhat into the background such that participants asked fewer questions about object labels and engaged more in non-task-related talk. As *truth-functional denial* on the participants' side typically occurs when Deechee answers wrongly to object-related questions, fewer such questions would lead to fewer occasions for Deechee to get it wrong and therefore fewer occasions for participants to correct it via the use of *truth-functional denial.* Another possibility for fewer *truth-*

functional denials on the participants' side is the presence of *self-prohibition* on Deechee's side in the prohibition experiment. Sometimes Deechee answered a participant's object-directed question with a word previously used by participants in the prohibition task such as *can't*. In this case participants typically did not take this to be an object label, but rather a case of self-prohibition. In such cases participants typically did not react with *truth-functional denial*. They seemed to understand that what Deechee said could not possibly be interpreted as object label and reacted by saying that it is ok for Deechee to take this object, that it is not forbidden any more, or other utterances of this kind. In the end only a complete analysis of all utterances could provide an answer to this question. Our analytical focus on negative utterances here prevents us from answering this question in a satisfactory manner.

Relating Negation Words to Negation Types

In this section we will look at the relationship between negation words and negation types. The question of interest with regards to acquisition is if certain negation words occur more frequently as part of certain negation types as compared to other negation types.

Within the word-level analysis (section 5.2) we showed that within both of our experiments participants produced various negation words, *no* being the kingpin amongst them both in terms of frequency and in terms of prosodic saliency. Furthermore we demonstrated that this relatively high production frequency of negation words contrasted to the way participants spoke in Saunders' experiment where they hardly produced any negative words. We subsequently assumed that it was the emotional displays and the motivationally congruent behaviours of the robot that elicited the participants' production of negation words as opposed to some other minor differences between our and Saunders' experiment such as the duration of the sessions. A major analytical drawback of the word-level is the

constraint to rather coarse-grained explanations that only refer to general differences between the two experiments. This level does not provide any explanation in terms of what our participants actually did, i.e. what conversational functions they performed, when producing these negation words. It is lacking any reference to the functional or pragmatic level of the utterances produced by our participants.

To fill this explanatory gap the pragmatic analysis so far showed the high prevalence of five of our 19 negation types in both experiments: *negative intent interpretations*, *negative motivational questions*, *truth-functional denials*, and, in case of the prohibition scenario, linguistic *prohibition* and *disallowance*. The high frequency of *negative intent interpretations* and *negative motivational questions* confirms the assumption from the word level: it is indeed the emotional displays and emotionally congruent behaviours which led our participants to produce these kind of utterances. If the robot does not display any emotion, or if it is not perceived to do so, participants evidently do not engage in linguistic interpretations of these expressions nor do they pose questions about its motivational state. Why a lack of emotional display simultaneously seems to suppress the production of *truth-functional denials*, which at least on the surface don't seem to be linked to emotional displays, is an open question.

Despite our knowledge about the frequencies of certain negation words - word level - and the frequencies of certain negation types - pragmatic level - up to this point we still don't know how precisely the two levels relate to each other. Of course we do know that any instance of some negation type contains some negation word, this is precisely how we constructed our negation types. However, we don't know if instances of certain negation types typically 'come along' with particular negation words. For this reason this section analyses the frequencies of negation words within the stated negation types. We will see that not all of the five highly-frequent negation types 'produce' the same amount of each

kind of negation word.

Of particular interest is *no* - typically the first negation word to emerge in child language development. If our data indicates that, firstly, certain frequently employed negation types[40] contain more *no*'s than others, and, secondly, that these *no*'s, when produced within instances of these types, are produced with at least average salience rate, we can conclude from which type of negation the majority of *no*'s in the robot's embodied lexicon ultimately originate.

Table 5.46 shows the total counts and percentages of negation words broken down by their occurrence within the most frequently observed negation types across both experiments. Furthermore, the table shows the salience rates of the selected negation words for the 5 most frequent negation types.

We have already seen within the analysis on the word level (section 5.2) that from our complete set of negation words produced by participants within our two experiments only four were highly frequent: *no*, *don't*, *not*, and, only within the prohibition scenario, *can't*. This observation can be seen replicated in table 5.46. Yet by means of grouping these negation words by their associated negation types this table shows us something that was not visible on the word level: the negation words produced by our participants are by no means equally distributed across the various negation types.

We already know from the previous pragmatic analysis that *can't* was exclusively used by participants within utterances that were either instances of *prohibitions* or *disallowances*. Yet, here we see that *no* was used nearly twice as often within prohibitive utterances as compared to *can't*. This comes as no surprise to the coders: *no* and *can't*

[40]Correctly we would have to say "frequently employed instances/utterances of a certain negation type". As this is a rather unwieldy expression, we will in the following speak as if utterances were types (thereby flattening the type-token distinction) as in "negation words produced as part of type xy". In all of these cases we really mean that negation words are produced as part of utterances which can be classified as being instances of certain negation types, and we textually conflate these relationships just in order not to discourage the reader with technically correct but overly bulky wordings.

often occurred simultaneously in the same utterance such as "No, you can't have that", and a singular *no* can work as a perfectly complete prohibition on its own. Importantly, as can be read from subtable (b), *no* also has a higher saliency rate as compared to *can't* within prohibitions. Yet, it would be premature to conclude from this that *no*'s, when employed within *prohibitions*, were, on average, prosodically more emphasized than *can't*s[41]. The reason for this is a second factor independent from prosodic emphasis which impacts on our implementation of prosodic saliency: utterance length. In a single-word utterance the one constitutive word will be extracted as the salient one no matter how weak the prosodic emphasis is. It therefore may come as no surprise that *no* has a higher salience rate as compared to *can't* within prohibitions. The obvious explanation for this being the case is that *no* is often produced on its own whereas *can't* is typically embedded in a larger utterance of length > 1. The only exception from this 'rule' would be if a speaker stops before and after his or her production of *can't*. This is possible, spoken language is often grammatically incomplete, but not very likely. We will shed some light on the issue of the explanatory tradeoff between utterance length and prosodic emphasis with regards to prosodic saliency in the next section.

A fairer comparison in terms of the saliency rate of *can't* is thus one that is performed between equals, i.e. between words that in the vast majority of cases don't come on their own. *Not* is one of these words that can hardly be produced solitarily in most variants of English[42]. And when comparing the different salience rates of *not* with the ones of *can't*,

[41] We use here *prosodically salient* to mean, that the word was chosen by our salience detection algorithm as the most salient word of the utterance. If an utterance consists of only one word it is therefore automatically and trivially prosodically salient. Contrastingly we use *prosodically emphasised* to refer to the acoustical properties of the production of the word. This means that the word was acoustically emphasized by the speaker through either higher pitch, higher energy, or both.

[42] To the knowledge of the author there exist some dialects of English which elevate *not* in its production frequency and grammatical role somewhat to the status of *no* such that it might be not uncommon to produce it solitarily in these dialects. Yet, British, American, and Nigerian English, which were employed by our participants do not form part of the dialects where this is the case.

we can see that *can't* fares rather well: its salience rate associated with *prohibition* (39.7%) is higher than any salience rate of *not*.

With respect to *no* it is interesting to see that within our five most frequent types, it typically has the highest share in terms of production of all negation words. The only exception is its share within *negative intent interpretations*: there it is second to *don't*. Interestingly we find the highest absolute as well as relative number of productions of *no* within utterances classified as *truth-functional denials*: Close to 70% of all negation words produced within this type are *no*'s. This could lead us to believe that this negation type is clearly the major source of *no*'s for the robot. However, if we take into account the salience rate of *no* within this type we can see that this is not the case: due to a comparatively low salience rate of *no* within productions of this type (34.6%), the absolute number of prosodically salient *no*'s is far lower as compared to the number of its salient productions within *negative intent interpretations* and *negative motivational questions*. There the salience rate of *no* is up to twice as large (79.6% in the case of *negative motivational questions*). This means that despite the overall larger number of productions of *no* within *truth-functional denials*, the majority of *no*'s within the robot's lexicons[43] derive from the two frequent motivational types. Added up, these two types are responsible for 262 instances of *no* within the robot's lexicons, as opposed to 73 provided by *truth-functional denials* - less than a third of the amount provided by the motivational types. The number of *no*'s provided by *prohibitions* seems on the first glance comparatively low. But if we take into account that *prohibitions* were only employed by participants during three out of the 5+5 sessions, its share of *no*'s could be expected to be the highest of all when multiplying the absolute

[43]We shall remind the reader at this point that the given absolute numbers are accumulated, i.e. they come about by lumping together the speech of all participants and all sessions. During runtime this is never the case, i.e. the robot only 'runs' with one of the 20 lexicons - the one based on the speech of the respective participant. Thus, in order to get a feel for the actual numbers of instances of words such as *no* in the robot's active lexicon, one has to divide the presented numbers for salient words by 20.

numbers by the factor 5/3 in order to compensate for the 'session-number disadvantage'.

The obvious danger in using accumulated measures such as the one used in table 5.46, i.e. summing up the productions of all participants into single cross-participant numbers, is that single very talkative participants could skew the measures such that the ensuing numbers do not represent the "average" participant any more. For this reason we statistically verified the validity of these production and salience rates by calculating the production rates for all words-type-participant combinations separately and subsequently employed ANOVAs and, in a few cases, T-tests on this data. The results, i.e. the true means and standard deviations on the one hand and an estimate of the statistical significance of the apparent differences in both production and salience rates on the other hand are presented in table 5.47.

Table 5.46: ***Negation words within most frequent negation types.*** (Top) *Listed are the absolute frequencies of negation words grouped by negation types as produced by all participants in all sessions within both experiments. The percentages in brackets give the share of the respective word relative to all negative words produced within the respective type.* (Bottom) *Listed are the number of salient words for each combination of negation word and type for the most frequently produced types and words. The percentages in brackets give the share of salient productions relative to the total number of productions of the respective word-type combination.*

(a) *All negation words by negation type: absolute and relative frequencies*

Type	neg. intent interpret.	neg. mot. question	truth-func. denial	prohibition	disallowance	truth-func. negation	neg. tag question
Word							
no	174 (39.2%)	191 (46.2%)	212 (67.9%)	129 (52.9%)	39 (45.3%)	14 (21.2%)	1 (1%)
not	59 (13.3%)	47 (11.4%)	93 (29.8%)	40 (16.4%)	27 (31.4%)	30 (45.5%)	0
don't	201 (45.3%)	164 (39.7%)	1 (0.3%)	2 (0.8%)	2 (2.3%)	2 (3%)	58 (58.6%)
isn't	0	9 (2.2%)	4 (1.3%)	0	0	0	18 (18.2%)
can't	0	0	0	68 (27.9%)	16 (18.6%)	1 (1.5%)	3 (3%)
haven't	0	0	1 (0.3%)	0	0	7 (10.6%)	1 (1%)
wasn't	0	0	1 (0.3%)	0	0	0	0
cannot	0	0	0	1 (0.4%)	1 (1.2%)	0	0
neither	0	0	0	0	1 (1.2%)	0	0
didn't	7 (1.6%)	1 (0.2%)	0	0	0	5 (7.6%)	12 (12.1%)
doesn't	1 (0.2%)	0	0	0	0	5 (7.6%)	3 (3%)
hasn't	0	0	0	0	0	1 (1.5%)	2 (2%)
weren't	0	0	0	0	0	0	1 (1%)
won't	2 (0.5%)	1 (0.2%)	0	0	0	0	0
mustn't	0	0	0	4 (1.6%)	0	1 (1.5%)	0

(b) *Salient negation words: absolute frequencies and percentage relative to total number of respective word*

Type	neg. intent interpret.	neg. mot. question	truth-func. denial	prohibition	disallowance
Word					
no	110 (62.9%)	152 (79.6%)	73 (34.6%)	85 (65.9%)	12 (30.8%)
not	17 (28.8%)	8 (17.0%)	7 (7.5%)	7 (17.5%)	4 (14.8%)
don't	39 (19.3%)	24 (14.7%)	0	1 (50.0%)	1 (50.0%)
can't	0	0	0	27 (39.7%)	8 (50.0%)

Table 5.47: ***Statistical analyses of relative frequencies of (salient) negation words within most frequent negation types.*** Subtables (a) and (b)*: ANOVAs comparing relative frequencies of the respective negation words across the stated negation types based on percentages in table B.25. Percentages of word-type combinations with less than the given number of instances were excluded from the analyses (cf. table B.25).* Subtable (c)*: ANOVAs and T-tests comparing salience rates of the respective negation words across negation types. T-Tests were employed instead of ANOVAs where only 2 groups were left due to insufficient data (cf. table B.27).* ***All numbers are percentages.*** *(brackets): standard deviation, ⋆: statistically significant with* $p <$ [p-value table]*, †: T-test instead of ANOVA due to two comparative groups with sufficient data only*

(a) *ANOVAs of relative type-word frequencies under exclusion of measurements based on* < 5 *instances*

Type	neg. intent interpret.	neg. mot. question	truth-func. denial	prohibition	disallowance	F	p
Word							
no	42 (14.3)	41 (27.4)	71 (16.7)	49 (24.8)	34 (42.2)	3.481⋆	0.014
not	14 (9.9)	11 (10.4)	29 (16.7)	17 (17.8)	45 (34.6)	5.046⋆	0.002
don't	45 (11.8)	48 (25.1)	0 (0)	0.9 (2.6)	1.7 (2.9)	28.54⋆	0.0001
can't	0 (0)	0 (0)	0 (0)	33 (26.2)	20 (29.2)	14.75⋆	0.0001

(b) *ANOVAs of relative word-type frequencies under exclusion of measurements based on* < 20 *instances*

Type	neg. intent interpret.	neg. mot. question	truth-func. denial	prohibition	disallowance	F	p
Word							
no	41 (13.5)	46 (20.6)	67 (20.6)	58 (19.9)	n/a	3.086⋆	0.046
not	15 (10.0)	13 (6.0)	33 (20.6)	16 (17.1)	n/a	3.112⋆	0.045
don't	45 (11.4)	41 (16.7)	0 (0)	0 (0)	n/a	30.57⋆	0.0001
can't	0 (0)	0 (0)	0 (0)	25 (20.1)	n/a	14.53⋆	0.0001

(c) *ANOVAs/T-tests of salience rates under exclusion of measurements based on* < 5 *instances*

Type	neg. intent interpret.	neg. mot. question	truth-func. denial	prohibition	disallowance	F/T	p
Word							
no	63.6 (16.6)	87.4 (21.7)	36.7 (20.4)	68.7 (21.5)	43.4 (51.3)	9.044⋆	0.0001
not	24.3 (32.1)	17.4 (15.3)	4.1 (9.0)	34.6 (49.2)	25.6 (31.0)	1.356	0.27
don't	15.4 (11.9)	13.7 (15.9)	n/a	n/a	n/a	0.328†	0.255
can't	n/a	n/a	n/a	31.7 (26.4)	62.5 (53.0)	1.292†	0.228

As can be seen in subtables (a) and (b), the differences in the means of production rates are all statistically significant. When we compare the means of percentages given in these two tables with the percentages based on the accumulated data in table 5.46 (*accumulated percentages*) we can see that the latter are not too far off the actual means. As stated in the captions of these subtables we excluded those percentages from the calculation of the mean which were based on utterance counts less than the stated threshold: 5 in the case of subtable (a) (criterion 1), and 20 in the case of subtable (b) (criterion 2). This was done in order to prevent a distortion of the total mean by individual (per participant, type, and word) percentages based on a very small number of utterances. To illustrate this problem, an example shall be in order.

Participant *P10* produced only a single utterance of the type *negative intent interpretation*, which contained a *no* (cf. table B.24). Subsequently the share of *no*'s relative to all negative words produced within this type by *P10* is 100%. This percentage can hardly be seen as representative for the share of a particular negation word relative to all negation words produced within this type by this participant, as the percentage is based on a single data point. For this reason this percentage was excluded from the calculation of the means under both exclusion criterions (marked *n/a* in tables B.25 and B.26 in the appendix). Thus, in order to be included in the calculation of the mean percentage, a participant must have produced at least 5 resp. 20 utterances of this type. If his or her productions fell short of these thresholds the counts for all negation words of this type for this participant were excluded from the respective analysis.

The underlying data basis for the calculation of these percentages is displayed in the appendix in the table B.24, the tables with the actual percentages and excluded values (based on latter table) are appended as well (tables B.25 and B.26). For more explanations on how the percentages were calculated, including exemplary calculations we refer to

section B.6 in the appendix.

As can be seen from the two mentioned subtables, the differences between the (statistically) calculated mean percentages or production rates and the accumulated percentages are typically in the range of up to 4%. Notable exceptions are the percentage for *don't* within *negative motivational questions* (approx. +8% in statistical mean under criterion 1, but not under criterion 2), the percentage for *no* within *prohibitions* (approx. +5% in-statistical mean under criterion 2, but not under criterion 1), and the percentage for *can't* within *prohibitions* (approx. +5% in statistical mean under criterion 1, but not under criterion 2).

Yet, as can be seen from subtable (c), only the difference between the means of the saliency rates for *no* are statistically significant. Our assertions that most *no*'s, that made it into the robot's active lexicon, originated to a large extent from productions of the two motivational negation types, *negative intent interpretations* and *negative motivational questions*, and, in case of the prohibition scenario, from productions of linguistic *prohibition* remains valid.

Our assertion that *can't*, when produced within *prohibitions*, had a higher salience rate than all productions of *not*, is slightly relativised. The salience rate of *not* within productions of *disallowance* is according to the ANOVA under exclusion criterion 1 34.6%, yet the difference of the means of production rates of *not* across the various types is not statistically significant. Therefore we cannot be sure how much we can trust the given means. The data basis was too small to perform an ANOVA under exclusion criterion 2. Nevertheless we may weaken our assertion to say that the mean salience rate of *can't* is higher than the mean salience rate of *not* for all negation types but one: the salience rate of *not* within prohibitions. When produced within *disallowances* its salience rate is even higher.

Interim Summary

The pragmatic analysis of negative utterances produced within the rejection experiment has shown that the vast majority of these utterances fall into one of three negation types: *negative intent interpretations*, *negative motivational questions*, and *truth-functional denials*. In which of these three types of negation a particular participant engaged most varied across participants. We thus can conclude that our expectations were realized: Firstly, participants of this experiment did engage in linguistic interpretations of the motivational or emotional state of the robot, and they posed questions with regard to these states to the robot. Secondly, instances of both of these two negation types frequently contain negative words as in "No, you don't like that" - this utterance could fall into either of these two categories depending on its intonational contour and conversational properties such as the presence or non-presence of a subsequent pause.

Within the prohibition experiment a new negation type took over the top-rank for the most frequently expressed: linguistic *prohibition*. A second new type emerged, *disallowance*, albeit the latter was less frequently produced as compared to *prohibitions*, *negative intent interpretation*, *negative motivational questions*, and *truth-functional denial*.

The analysis of saliency rates applied to the rejection experiment showed that negation words when expressed within the two motivational types, *negative intent interpretation* and *negative motivational questions* have an approximately 10% higher saliency rate as compared to them being expressed within *truth-functional denials*. Applying the same analysis to the prohibition experiment yielded a similar result, although the difference in saliency rates between the two motivational types and *truth-functional denial* is less stark. However, it appeared that negation words expressed within linguistic *prohibitions* reached higher saliency rates than when expressed in any other negation type.

When comparing participants' speech of both experimental experiments against each

other the first, now obvious, difference is that participants within the prohibition experiment engaged in linguistic *prohibitions* and *disallowances* whereas participants in the rejection experiment did not.

Slightly less obvious, participants from the prohibition experiment showed a strong tendency to engage less in *negative intent interpretations* as compared to their peers in the rejection experiment within the first three sessions - these are the sessions during which the prohibitive task was present or active. Furthermore there is a (less strong) tendency for them to also engage less in *negative motivational questions* as compared to their peers during the same first three sessions. Finally participants from the prohibition experiment engage significantly less in *truth-functional denials* during the last two sessions as compared to their 'rejective' peers during these sessions. We cannot explain on this analytical level why this is the case.

Finally, when looking in more detail at the production rates of particular negation words, *no*, *don't*, *not*, and *can't*, within the frequent negation types, we have shown that they are not equally distributed across productions of these types. Albeit *prohibitions* and *disallowances* were responsible for all the *can't*s within the corpus, these types are also responsible for nearly twice as many *no*'s. Due to these *no*'s having the second-highest saliency rate when being expressed within *prohibitions*, this type is responsible for more *no*'s in the PCS of the first three sessions than the other two motivational types, *negative intent interpretations* and *negative motivational questions*, taken together.

Despite *truth-functional denials* containing overall the highest absolute production rate of *no*'s, their saliency rate within productions of this type is only about half as big as compared to their saliency rate within the motivational types. This allows for *truth-functional denials* to contribute only ∼17% of all *no*'s to the total amount of *no*'s 'generated' by the

5 most frequent types[44]. In contrast, *negative intent interpretations*, *negative motivational questions*, and *prohibition* contribute ∼25%, ∼35%, and ∼20% to this set of *no*'s. Albeit the said percentage of *prohibitions* (∼20%) might not strike the reader as considerably larger than the one of *truth-functional denials* (∼17%), we would like to emphasise the fact that *prohibitions* are only expressed during the first 3 sessions within the prohibition scenario whereas instances of the other aforementioned types are expressed during all 5+5 sessions of both experiments[45], thus rendering *prohibitions* much more effective in terms of their contribution of *no*'s to the global corpus.

On Prosodic Saliency

An important notion, theoretically as well as in terms of our implementation, is the notion of a word being prosodically salient. The operationalisation of prosodic saliency within Saunders' word extraction mechanism is such, that it is assumed that every utterance contains at least one prosodically salient word. The extracted word subsequently becomes part of the embodied lexicon of the robot. We have seen in the previous section that the prosodic salience of various negation words varies between different negation types. Especially interesting for our research question regarding the origins of negation is the circumstance that *no* shows within motivational negation types a higher salience rate as compared to it being expressed within *truth-functional denials*. We could take this as an indication that *no* might be initially a primarily motivational entity for the language learner precisely because it is prosodically emphasized when being used in emotional kinds

[44] 5 most frequent types: *truth-functional denial*, *prohibition*, *motivation-dependent question*, *negative intent interpretation*, *disallowance*.

[45] This number is not entirely correct for *truth-functional denials*. That type is typically expressed by participants as soon as the robot labels an object incorrectly. As the robot only speaks from the 2nd session onwards, possibly all but certainly the majority of *truth-functional denials* are expressed within the 2nd to 5th session. Thus the number of sessions during which this type is most likely to be expressed is 4+4=8, which are still more than twice as many sessions as compared to those within which *prohibitions* are 'active'.

of negation by the teacher.

In order to shed some more light on this issue we will look at the second, competing factor with regards to prosodic salience: utterance length. Remember that our word extraction algorithm always extracts the prosodically most salient word from any utterance. If an utterance consists of only a single isolated word, as is the case in a simple "No!", this word will automatically, and trivially, be the prosodically most salient one. For these cases we are not able to say, on the basis of our available data set, if this word was actually prosodically emphasised or if it was salient by virtue of being the only word of the utterance.

Figures 5.7 and 5.8 show boxplots which visualize the distribution of utterance lengths for all productions of the displayed word-type combinations where the stated negation word is salient. These figures are contrasted with boxplots of the utterance length of *all* productions of the same word-type combinations. Naturally the former are a subset of the latter.

We can observe that all *no*'s, independent within which negation type they are produced, come frequently as self-containing utterances (utterance length (ul) = 1), whereas this is only rarely the case for any of the other negation words. If we compare the general length distribution of all utterances (red, leftmost boxes) with the various length distributions of utterances containing *no*, we can see that the latter are generally lower with one exception: the distribution of *truth-functional denials*. There, if all utterances of this type and word are considered, the median *ul* equals 3 and 50% of all utterances have an utterance length between 1 and 5[46]. If we compare this to the length distribution of all utterances, independent of type (red, left-most box) we can see that both distributions look very similar. Yet, if we compare this to the length distribution of the subset of *truth-functional denials* whose *no* is prosodically salient, the picture changes drastically: there

[46]We will abbreviate this in the following with the notation "median(ul) = 3 && $1 \leq 50\% \leq 5$"

$median(ul) = 1$ and 50% of all utterances consist of single *no*'s[47]. This presents an interesting contrast to salient productions of *not* within *truth-functional denials* as the latter have a median *ul* of 4 and 50% of these utterances are between 4 and 5 words long. We mentioned in the previous section that *not* hardly ever 'comes along alone' based on our coding experience. Here we have the statistical confirmation of this very assertion.

When comparing the productions of *no* within *truth-functional denials* with its production within the motivational types, it appears that the former are slightly longer than the latter (*truth-functional denials*: $median(ul) = 3$ && $1 \leq 50\% \leq 5$, *negative intent interpretations*: $median(ul) = 1$ && $1 \leq 50\% \leq 4$, *negative motivational questions*: $median(ul) = 1$ && $1 \leq 50\% \leq 2$). When focusing on the subset of salient productions of *no* across the same types (blue, dotted boxes), the difference in length distribution seems to disappear (*truth-functional denials (salient)*: $median(ul) = 1$ && $50\% = 1$, *negative intent interpretations (salient)*: $median(ul) = 1$ && $50\% = 1$[48], *negative motivational questions (salient)*: $median(ul) = 1$ && $50\% = 1$[49]).

This indicates that the frequently salient *no*'s within the motivational negation types as well as within *truth-functional denial* were not necessarily salient because they were prosodically emphasised by the speakers, but rather because they were produced in isolation. Naturally, by just looking at distributions of utterance lengths we can not definitively say that these *no*'s, or even a majority of them, were not prosodically emphasised. But the fact that most prosodically salient *no*'s were produced in isolation hints towards utterance length being an important factor for them being salient. This is especially true for *truth-functional denials* where the prosodically salient *no*'s are significantly more often found in utterances of length 1 as opposed to all productions of *no* within utterances of this type.

[47] more precisely: ∼78% of all utterances have length 1 (taken from data as precise percentage is not readable from boxplot).

[48] more precisely: ∼83% of all utterances have length 1.

[49] more precisely: ∼91% of all utterances have length 1

When we finally focus on productions of *no* within *prohibitions*, we can discern a similar trend. While the median utterance length of all, i.e. salient and non-salient, productions of *no* is 2 and 50% of the utterances are between 1 and 5 words long, the median utterance length of only salient productions is 1 and 50% of these productions are between 1 and 2 words long[50].

Summary Based on the given analysis of utterance lengths we can now say with certainty that *no*'s, independent of the kind of negation within which they were expressed, were part of relatively short utterances compared to the average utterance produced by our participants. Many of these productions even consisted solely of these *no*'s and this observation holds true for all frequently observed negation types. This may be seen in contrast to other negation words, where this is not the case: utterances containing *not*, *don't*, *can't* etc. are, on average, longer compared to utterances containing (or solely consisting of) *no*. Furthermore we cannot conclude that the majority of prosodically salient productions of *no*'s have been prosodically emphasised but they may have been frequently salient by virtue of constituting single-word utterances. However, we can not exclude with certainty the possibility that they may have been prosodically emphasized as we did not analyse prosody directly. Only a direct prosodic comparison between these negation words and non-negation words could yield any definite conclusion.

[50] more precisely: ∼72% of all utterances have length 1.

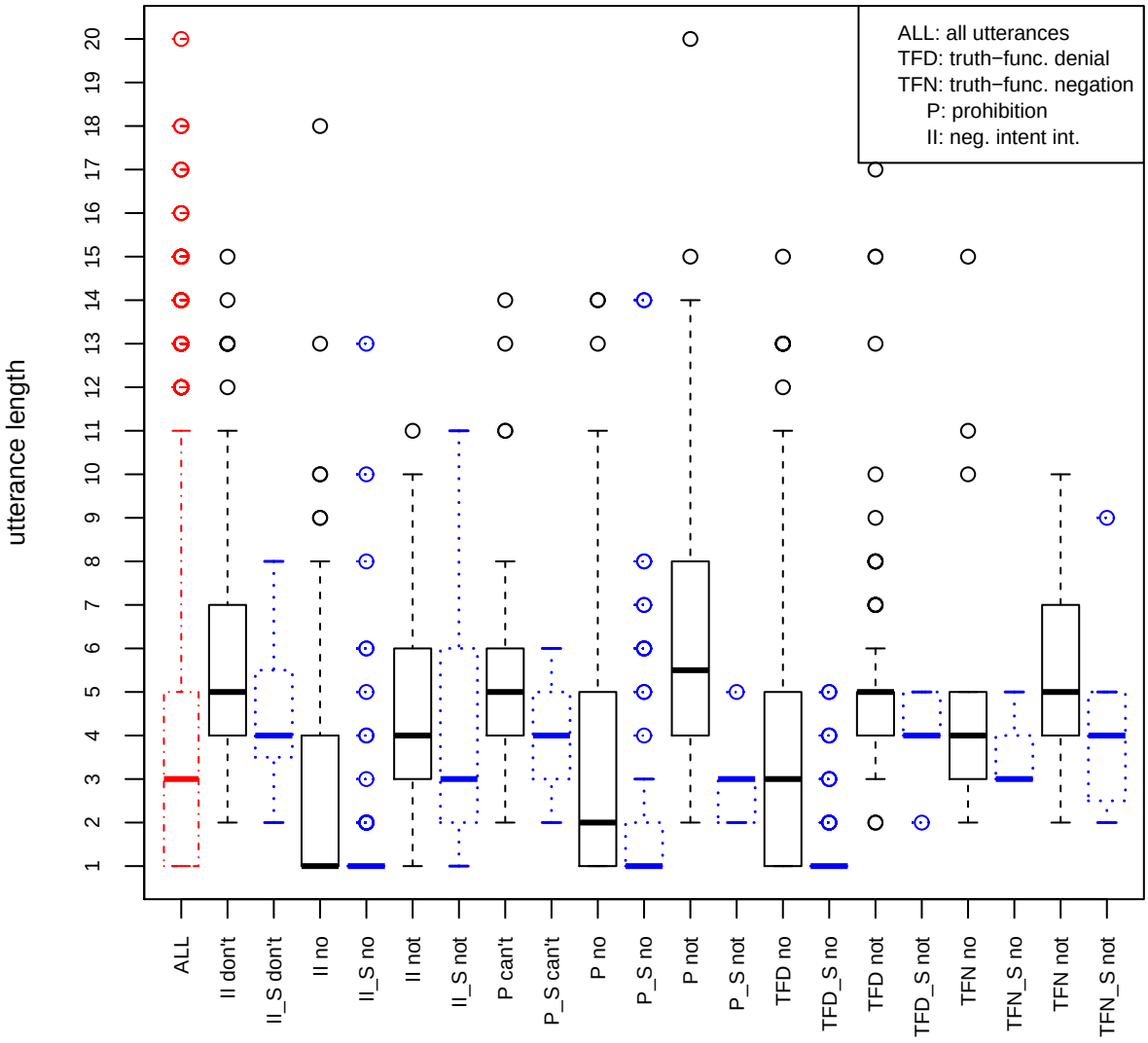


Figure 5.7: ***Utterance lengths of utterances containing the most frequent negation words, grouped by negation type (1).*** *Displayed are box plots of the utterance lengths of the most frequently produced negation words of the most frequent negation types, grouped by negation type. For comparison the distribution of utterances lengths of all utterances is displayed as well on the very left. The median is displayed as the bold bar in the middle of the box, the lower and upper boundaries of the box are the 1. and 3. quartile respectively. The whiskers extend down- and upwards up to 1.5*IQR (inter-quartile range). Black, solid boxplots: The underlying data is based on all instances of the stated words. Blue, dotted boxplots: The underlying data is based on only those instances of the stated words where this word was identified as being prosodically salient. Red, dot-dashed boxplot: underlying data are utterances lengths of all utterances.*

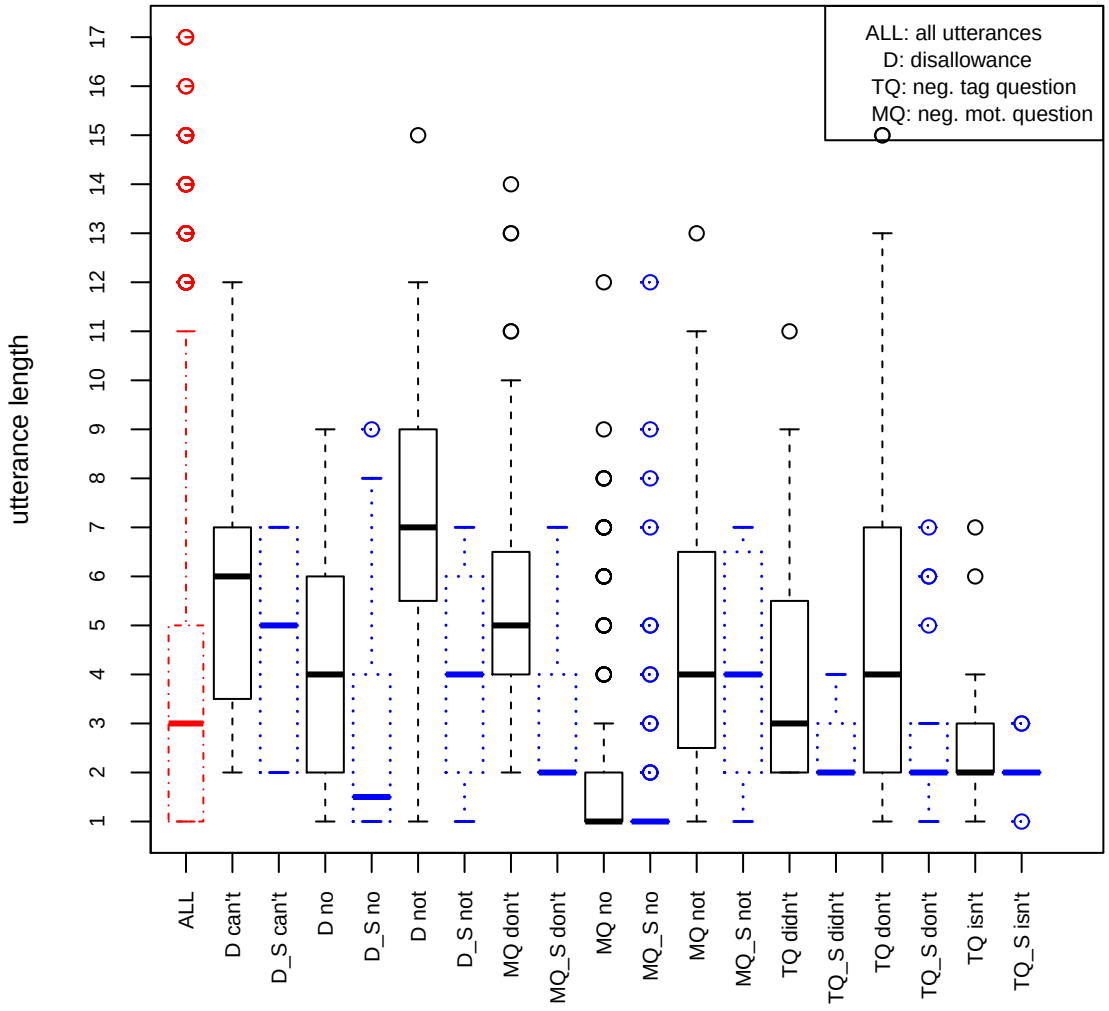


Figure 5.8: ***Utterance lengths of utterances containing the most frequent negation words, grouped by negation type [2].*** *Displayed are box plots of the utterance lengths of the most frequently produced negation words of the most frequent negation types, grouped by negation type. For comparison the distribution of utterances lengths of all utterances is displayed as well on the very left. The median is displayed as the bold bar in the middle of the box, the lower and upper boundaries of the box are the 1. and 3. quartile respectively. The whiskers extend down- and upwards up to 1.5*IQR (inter-quartile range). Black, solid boxplots: The underlying data is based on all instances of the stated words. Blue, dotted boxplots: The underlying data is based on only those instances of the stated words where this word was identified as being prosodically salient. Red, dot-dashed boxplot: underlying data are utterances lengths of all utterances.*

5.4 Robot: Evaluation of Acquisition

In this section we will focus on the robot's part of the human-robot dialogue and evaluate its success in acquiring the ability to engage in negation. The analysis is based on the first coder's decision whether Deechee's negative utterances were adequate or felicitous in the respective situations. By doing so we treat the robot as a black box, i.e. we consciously don't pay attention to any of its internal workings, but rather make the judgment as an external observer of the conversation without any formal criterion for felicity, similar to the way that one may judge the linguistic capabilities of a child.

Unfortunately, as we have seen in section 5.3.5, the intercoder agreement on the felicity of Deechee's utterances was very low. The most frequent reason for this disagreement, after exclusion of pragmatic negatives, turned out to be disagreement on Deechee's intention. In our particular context this means that the two coders did not agree if Deechee really did or did not want a particular object. This will have an especially large impact with regards to decisions on felicity of utterances, which would in a subsequent coding stage be classed as instances of one of the motivational negation types.

Yet, as this is the only available measure for judging the robot's performance similar to how we may judge children's linguistic performance, we will use it despite the low intercoder agreement. The good news in this context is that there is no indication that the first coder, whose decisions form the basis of this analysis, did judge the robot's linguistic performance more favourably as compared to the second coder. Table 5.48 lists the two coders' felicity rates on the full coding set and the three subsets where either *P12*'s utterances, utterances containing pragmatic negatives, or both of the former are excluded. As can be seen there, on 2 out of these 4 sets the 1st coder judged more of the robot's negative utterances as felicitous compared to the 2nd coder and on the other 2 the contrary holds.

Table 5.48: ***Felicity rates of the two coders:*** full*: full coding set, no exclusions,* -prag.*: coding set where utterances containing pragmatic negatives are removed,* -P12*: coding set where* P12*'s utterances are excluded,* -P12 && -prag.*: coding set where both* P12*'s and pragmatic negatives are excluded.*

coder	full	- prag.	- P12	- P12 && -prag.
c1	53.91	46.94	54.88	48.57
c2	51.94	48.98	54.35	55.71

Pragmatic negatives were not considered for the following analysis due to their detrimental effect upon the intercoder agreement (cf. section 5.3.5). For the coding set without utterances containing pragmatic negatives the 1st coder judged less utterances to be felicitous (46.94%) as compared to the 2nd coder (48.98%). We expect for this reason the percentages pertaining to the felicity of the robot's utterances, presented in the following, to be under- rather than overestimates.

Tables 5.49 and 5.50 give an overview of the negation types produced by Deechee within the respective experiments as identified by the 1st coder. Within the rejection experiment the robot did produce negation words in its conversation with 7 out of the 10 participants but not with the participants *P01*, *P06*, and *P10*. Furthermore we can see that even for those participants where Deechee did produce negative words, it did not necessarily produce these within each session. In the conversation with *P04*, for example, Deechee produced negative words during the third and fifth, but not during the second and fourth session.

This can happen because Deechee's productions depend entirely on the content of its embodied lexicon. The latter is updated in between sessions, and the update consists of the salient words which the respective participant produced in the session before. The words prosodic saliency for their part depend on how and in which lexical context the participants produced them in terms of utterance lengths and prosodic emphasis. Furthermore the update also depends on whether what the participants said was uttered in a temporally

congruent fashion. With *temporally congruent* we mean that an object is labeled while it is being presented, not before or too long after the presentation. The same applies to *negative intent interpretations*, *negative motivational questions* and *prohibitions*: from an algorithmic perspective and in case of the two motivational types they should be produced while Deechee was in the respective motivational state, not before it entered that state and not long after it left that state. Similarly from the viewpoint of the architectural design we expected participants to express linguistic *prohibition* while they were physically prohibiting the robot, not before or too long after that.

Albeit our acquisition architecture has certain time buffers and other 'coping mechanisms' in place to handle slight delays in a participant's production relative to the physical state of affairs, all of these mechanisms are very limited. The principal assumption backing the design of the acquisition algorithm is that linguistic production happens roughly simultaneously to the state of affairs that is referred to within the 'content' of the production. Overly severe violations of this constraint will have a detrimental impact on the robot's learning success. In such cases bad exemplars enter the lexicon where the respective word is associated with non-fitting sensorimotor-motivational data. Subsequently it can happen that what in a previous session seemed to be a 'well-understood' word, becomes dissociated from the correct referent due to the presence of 'bad', non-fitting exemplars. The respective word might then not be expressed any more in the following session or it might be expressed in the wrong sensorimotor-motivational context.

As compared to the rejection experiment we don't find any single participant-robot dyad within the prohibition experiment where Deechee did not produce any negation word in the course of all five sessions. Yet we do find participants where the robot hardly engaged in negation. With participant *P13* Deechee only uttered four negative words during session 4, and another single one during session 5. With *P14* Deechee did not use any negation in

the second and third session during which the prohibition task was in place, whereas with *P18* the opposite was the case: Deechee used three types of negation five times during the second session just to abandon any negative linguistic activity in all of the remaining three sessions.

Table 5.51 sums up all sessions displayed in the previous two tables in a more compact format for the purpose of comparing both experiments directly. We may call the absolute total stated in the bottom-right corner an accumulated measure. With respect to this measure every utterance has the same impact on the resulting felicity. As discussed in previous sections such measures are skewed towards sessions where the robot was more talkative, or more precisely, such sessions where the robot produced more negations. For this reason we complement this analysis with a statistical one where every robot-participant dyad is given the same weight within the total measure.

Table 5.49: ***Frequency of robot negation types and their felicity - Rejection experiment****. Listed are the negation types which the robot engaged in for all participants and sessions. Given for each type are its frequency (*cnt*) and the percentage of cases in which the robots engagement in this type was judged as felicitous. The last line of each session contains the accumulated measures of frequency and percentage of felicity across all observed types. As the robot did not speak during the 1. session the listing starts with session nr. 2. The following abbreviations are used for the negation types:* TD*: truth-func. denial,* MD*: mot.-dep. denial,* A*: neg. agreement,* R*: rejection,* I*: neg. imperative,* E*: mot. dep. exclamation.*

		P01		P04		P05		P06		P07		P08		P09		P10		P11		P12	
sid	**type**	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel
s2	TD	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	6	0
	A	0	n/a	0	n/a	7	100	0	n/a	1	100	0	n/a	3	100	0	n/a	0	n/a	0	n/a
	R	0	n/a	0	n/a	8	25	0	n/a	0	n/a	6	16.67	6	100	0	n/a	0	n/a	6	33.33
	MD	0	n/a	0	n/a	17	52.94	0	n/a	4	100	6	83.33	3	66.67	0	n/a	0	n/a	3	66.67
	I	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	11	81.82	0	n/a	0	n/a	0	n/a
	all	0	n/a	0	n/a	32	56.25	0	n/a	5	100	12	50	23	86.96	0	n/a	0	n/a	15	26.67
s3	TD	0	n/a	4	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	3	100
	E	0	n/a	0	n/a	0	n/a	0	n/a	1	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	A	0	n/a	0	n/a	0	n/a	0	n/a	2	50	0	n/a	1	100	0	n/a	2	100	0	n/a
	R	0	n/a	1	100	3	0	0	n/a	1	0	2	50	1	100	0	n/a	6	100	7	71.43
	MD	0	n/a	0	n/a	6	0	0	n/a	2	50	2	100	3	100	0	n/a	2	100	0	n/a
	I	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	4	100	0	n/a	0	n/a	0	n/a
	all	0	n/a	5	100	9	0	0	n/a	6	50	4	75	9	100	0	n/a	10	100	10	80
s4	E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	100	0	n/a	0	n/a	0	n/a	0	n/a
	A	0	n/a	0	n/a	0	n/a	0	n/a	3	100	1	100	0	n/a	0	n/a	0	n/a	0	n/a
	R	0	n/a	0	n/a	3	0	0	n/a	1	100	3	66.67	0	n/a	0	n/a	0	n/a	0	n/a
	MD	0	n/a	0	n/a	6	33.33	0	n/a	5	60	4	75	0	n/a	0	n/a	4	0	3	0
	I	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	5	100	0	n/a	0	n/a	0	n/a
	all	0	n/a	0	n/a	9	22.22	0	n/a	9	77.78	9	77.78	5	100	0	n/a	4	0	3	0
s5	TD	0	n/a	1	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	0	4	100
	E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	A	0	n/a	4	100	1	100	0	n/a	1	100	0	n/a	0	n/a	0	n/a	0	n/a	1	100
	R	0	n/a	2	100	3	100	0	n/a	2	100	0	n/a	1	100	0	n/a	10	50	3	100
	MD	0	n/a	0	n/a	1	100	0	n/a	4	75	1	100	1	100	0	n/a	4	50	0	n/a
	I	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	all	0	n/a	7	100	5	100	0	n/a	7	85.71	1	100	2	100	0	n/a	15	46.67	8	100

Table 5.50: ***Frequency of robot negation types and their felicity - Prohibition experiment.*** *For explanation of symbols please see previous table 5.49. Additional abbreviations:* SP*: self-prohibition,* PD*: perspective-dependent denial*

		P13		P14		P15		P16		P17		P18		P19		P20		P21		P22	
sid	**type**	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel
s2	TD	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	n/a	0	n/a	0	n/a
	MD	0	n/a	0	n/a	15	46.67	1	100	0	n/a	1	100	0	n/a	4	50	0	n/a	0	n/a
	PD	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	4	100	0	n/a	0	n/a	0	n/a
	R	0	n/a	0	n/a	5	40	0	n/a	3	100	2	100	0	n/a	3	100	3	66.67	0	n/a
	A	0	n/a	0	n/a	7	85.71	1	100	0	n/a	0	n/a	0	n/a	6	83.33	2	100	0	n/a
	I	0	n/a	0	n/a	0	n/a	1	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	SP	0	n/a	0	n/a	2	0	0	n/a	0	n/a	0	n/a	5	100	5	60	0	n/a	0	n/a
	all	0	n/a	0	n/a	29	51.72	3	100	3	100	3	100	9	100	19	68.42	5	80	0	n/a
s3	TD	0	n/a	0	n/a	0	n/a	0	n/a	2	50	0	n/a	0	n/a	0	n/a	10	30	1	100
	MD	0	n/a	0	n/a	6	50	0	n/a	3	100	0	n/a	0	n/a	0	n/a	3	0	0	n/a
	PD	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	0	0	n/a	0	n/a
	R	0	n/a	0	n/a	5	60	0	n/a	7	100	0	n/a	0	n/a	0	n/a	3	33.33	0	n/a
	A	0	n/a	0	n/a	4	75	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	6	66.67	3	100
	I	0	n/a	0	n/a	0	n/a	3	0	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	100
	SP	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	6	100	0	n/a	11	100	0	n/a
	all	0	n/a	0	n/a	15	60	3	0	12	91.67	0	n/a	6	100	1	0	33	57.58	5	100
s4	TD	4	0	2	0	1	0	1	0	4	0	0	n/a	0	n/a	1	0	0	n/a	2	0
	MD	0	n/a	0	n/a	8	37.5	5	0	12	50	0	n/a	0	n/a	6	16.67	5	80	0	n/a
	PD	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	3	33.33	0	n/a	0	n/a
	R	0	n/a	3	33.33	2	0	0	n/a	3	33.33	0	n/a	0	n/a	1	0	0	n/a	0	n/a
	A	0	n/a	0	n/a	1	100	3	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	100
	SP	0	n/a	0	n/a	0	n/a	1	0	0	n/a	0	n/a	3	100	2	0	4	0	0	n/a
	all	4	0	5	20	12	33.33	10	30	19	36.84	0	n/a	3	100	13	15.38	9	44.44	3	33.33
s5	TD	1	0	8	0	2	0	0	n/a	2	0	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	MD	0	n/a	1	0	9	22.22	1	0	13	23.08	0	n/a	0	n/a	2	0	4	75	0	n/a
	PD	0	n/a	1	100	0	n/a	0	n/a	1	0	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
	R	0	n/a	0	n/a	1	100	1	0	2	0	0	n/a	0	n/a	0	n/a	4	100	1	0
	E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	0	0	n/a	0	n/a
	SP	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	2	100	0	n/a	1	0	0	n/a
	all	1	0	10	10	12	25	2	0	18	16.67	0	n/a	2	100	3	0	9	77.78	1	0

Table 5.51: ***Accumulated frequencies for robot negation types and their felicity.*** *Displayed are the accumulated frequencies of the various negation types the robot engaged in across sessions and their felicities. The following abbreviations are used for the negation types:* TD*: truth-func. denial,* MD*: mot.-dep. denial,* A*: neg. agreement,* R*: rejection,* I*: neg. imperative,* E*: mot. dep. exclamation,* SP*: self-prohibition,* PD*: perspective-dependent denial.*

(a) *Rejection Experiment*

	P01		P04		P05		P06		P07		P08		P09		P10		P11		P12		*total*	
type	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	*cnt*	*%fel*
TD	0	n/a	5	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	0	13	53.85	*19*	*63.16*
E	0	n/a	0	n/a	0	n/a	0	n/a	1	100	1	100	0	n/a	0	n/a	0	n/a	0	n/a	*2*	*100*
A	0	n/a	4	100	8	100	0	n/a	7	85.71	1	100	4	100	0	n/a	2	100	1	100	*27*	*96.3*
R	0	n/a	3	100	17	29.41	0	n/a	4	75	11	36.36	8	100	0	n/a	16	68.75	16	62.5	*75*	*58.67*
MD	0	n/a	0	n/a	30	40	0	n/a	15	73.33	13	84.62	7	85.71	0	n/a	10	40	6	33.33	*81*	*56.79*
I	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	20	90	0	n/a	0	n/a	0	n/a	*20*	*90*
all	0	n/a	12	100	55	45.45	0	n/a	27	77.78	26	65.38	39	92.31	0	n/a	29	58.62	36	55.56	*224*	*66.07*

(b) *Prohibition Experiment*

	P13		P14		P15		P16		P17		P18		P19		P20		P21		P22		*total*	
type	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	*cnt*	*%fel*
TD	5	0	10	0	3	0	1	0	8	12.5	0	n/a	0	n/a	2	0	10	30	3	33.33	*42*	*11.9*
E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	0	0	n/a	2	100	*3*	*66.66*
A	0	n/a	0	n/a	12	83.33	4	100	0	n/a	0	n/a	0	n/a	6	83.33	8	75	3	100	*33*	*84.84*
R	0	n/a	3	33.33	13	46.15	1	0	15	73.33	2	100	0	n/a	4	75	10	70	1	0	*49*	*61.22*
MD	0	n/a	1	0	38	39.47	7	14.29	28	42.86	1	100	0	n/a	12	25	12	58.33	0	n/a	*99*	*39.39*
I	0	n/a	0	n/a	0	n/a	4	25	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	*4*	*25*
PD	0	n/a	1	100	0	n/a	0	n/a	1	0	0	n/a	4	100	4	25	0	n/a	0	n/a	*10*	*60*
SP	0	n/a	0	n/a	2	0	1	0	0	n/a	0	n/a	16	100	7	42.86	16	68.75	0	n/a	*42*	*71.43*
all	5	0	15	13.33	68	45.59	18	33.33	52	46.15	3	100	20	100	36	41.67	56	60.71	9	66.67	*282*	*50*

In order to pitch our two research hypotheses on the origins of negation against each other, hypothesis 1, the hypothesis that negation may originate in the linguistic interpretation of a child's motivational state by the caregiver (*intent interpretations*), was taken to be the baseline for both experiments. Hypothesis 2, the hypothesis that a toddler's use of negation is rooted in parental prohibition, on the other hand was modeled as optional treatment in the form of the prohibition task, present during the first three sessions of the prohibition experiment. We thus compared the sessions 4 and 5 of the two experiments for the respective rates of felicity of their respective utterances in order to evaluate the effect of the treatment. The result of this comparison is displayed in table 5.52.

Ostensibly the felicity rate of all negative utterances across all participants during these two last sessions is considerably lower within the prohibition experiment (30.15%) as compared to the rejection experiment (67.86%): less than every third negative utterance within the prohibition experiment was deemed felicitous, whereas this was the case for more than every second negative utterance within the rejection experiment. On the face of it these numbers indicate that the "prohibitive treatment" had a rather detrimental effect upon the robot's performance of negative linguistic acts. This result took us somewhat by surprise as the observations that we made during the execution of the experiment seemed to indicate that participants would quite reliably engage in linguistic *prohibition*, and, furthermore, that this negation type was a very reliable source for negation words. In order to somewhat alleviate this surprise we therefore conducted a separate analysis of some assumptions of our experimental design, which will be presented in the following section 5.5.

Table 5.52: ***Accumulated frequencies for sessions 4+5 for robot negation types and their felicity.*** *Displayed are the accumulated frequencies of the various negation types the robot engaged in during the last two sessions and their felicities. The following abbreviations are used for the negation types:* TD*: truth-func. denial,* MD*: mot.-dep. denial,* A*: neg. agreement,* R*: rejection,* I*: neg. imperative,* E*: mot. dep. exclamation,* SP*: self-prohibition,* PD*: perspective-dependent denial.*

(a) *Rejection Experiment*

	P01		P04		P05		P06		P07		P08		P09		P10		P11		P12		*total*	
type	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel
E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	100	0	n/a	0	n/a	0	n/a	0	n/a	*1*	*0*
TD	0	n/a	1	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	0	4	100	*6*	*83.33*
A	0	n/a	4	100	1	100	0	n/a	4	100	1	100	0	n/a	0	n/a	0	n/a	1	100	*11*	*100*
R	0	n/a	2	100	6	50	0	n/a	3	100	3	66.67	1	100	0	n/a	10	50	3	100	*28*	*67.86*
MD	0	n/a	0	n/a	7	42.86	0	n/a	9	66.67	5	80	1	100	0	n/a	8	25	3	0	*33*	*48.48*
I	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	5	100	0	n/a	0	n/a	0	n/a	*5*	*100*
all	0	n/a	7	100	14	50	0	n/a	16	81.25	10	80	7	100	0	n/a	19	36.84	11	72.73	*84*	*67.86*

(b) *Prohibition Experiment*

	P13		P14		P15		P16		P17		P18		P19		P20		P21		P22		*total*	
type	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel	cnt	%fel
TD	5	0	10	0	3	0	1	0	6	0	0	n/a	0	n/a	1	0	0	n/a	2	0	*28*	*0*
MD	0	0	1	0	17	29.41	6	0	25	36	0	n/a	0	n/a	8	12.5	9	77.78	0	n/a	*66*	*33.33*
PD	0	n/a	1	100	0	n/a	0	n/a	1	0	0	n/a	0	n/a	3	33.33	0	n/a	0	n/a	*5*	*40*
R	0	n/a	3	33.33	3	33.33	1	0	5	20	0	n/a	0	n/a	1	0	4	100	1	0	*18*	*38.89*
A	0	n/a	0	n/a	1	100	3	100	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	*4*	*100*
E	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a	1	0	0	n/a	1	100	*2*	*50*
SP	0	n/a	0	n/a	0	n/a	1	0	0	n/a	0	n/a	5	100	2	0	5	0	0	n/a	*13*	*38.46*
all	5	0	15	13.33	24	29.17	12	25	37	27.03	0	n/a	5	100	16	12.5	18	61.11	4	25	*136*	*30.15*

As mentioned in previous sections, basing our analysis on absolute numbers has the disadvantage that the total represents those human-robot dyads more, where Deechee was more (negatively) talkative, than those, where Deechee was rather taciturn with regard to the expression of negatives.

In order to remedy this concern, a t-test was performed on the two sets of felicity values of the participants in each experiment, independent of the negation type (bottom row "all" in table 5.52). Two t-tests were performed, firstly, on the unpruned data set[51] (criterion 1), and secondly, on those felicity values that are based on at least 10 utterances or more in order to delimit the impact of extreme outliers caused by an insufficient data basis (criterion 2).

Table 5.53: ***Statistical comparison of felicity rates between both experiments****: Given are the mean and standard deviation for the felicities of the robot's production of negation during the sessions 4 and 5 under two criteria:* ***Crit. 1****: data basis = felicity values of all participants but* P01*,* P06*,* P10*, and* P18*.* ***Crit. 2****: data basis as in crit. 1 plus additional exclusion of* P04*,* P09*,* P13*,* P19*, and* P22*.* $\star : p < 0.01$*,* $\dagger : p < 0.02$*.*

criterion	experiment	mean % felicity (std)	T
crit. 1	R	74.4 (23.8)	$3.0^{\star}$
	P	32.57 (30.31)	
crit. 2	R	64.16 (19.8)	$3.2^{\dagger}$
	P	28.02 (17.08)	

The results of the t-test, shown in table 5.53, confirm the result of the accumulated measure: In terms of felicity (or adequacy) of its negative utterances Deechee performed considerably worse within the last two sessions of the prohibition experiment as compared to the last two sessions of the rejection experiment. The difference in means is in both cases statistically significant.

[51] Naturally even here human-robot dyads where the robot did not produce any negation whatsoever, i.e. with participants *P01*,*P06*, *P10*, and *P18*, were disregarded (value *n/a*).

5.5 Human + Robot: Temporal Relationships

The result of the pragmatic analysis, i.e. the comparatively bad linguistic performance of Deechee in terms of felicity within the prohibition experiment came as a surprise to the author and seemed to run counter to our (unquantified) observations during the execution of the experiments. While we observed that *negative intent interpretations* were performed at times without the use of negation, for example by saying "oh, you look sad" instead of "no, you don't like it", we did not observe a single participant engaging in linguistic *prohibition* that would not involve some negation word.

For this reason we suspected that one of our algorithmic constraints had been violated overly much, the constraint that the 'referent', in our case the application of a corporal constraint (called *push* in the following), and linguistic production occur roughly simultaneously. We will refer to the latter in the following as *simultaneity constraint*.

5.5.1 Temporal relations between prohibitive action and linguistic prohibition

In order to specify more precisely which temporal relations between *pushes* and linguistic *prohibitions* and *disallowances* manifested during the prohibition scenario, we reconstructed the precise timings of utterances of both of the aforementioned types, subsequently termed *prohibition*$^{+}$, based on timed transcriptions, pragmatic codes, and the timings of the participant's corporal restraint of the robot. The latter were extracted from the readings of the robot's arm-pressure sensor from its log-files. We subsequently fused the different time-lines of these two sources into one consistent time line and plotted the temporal profile of *prohibition/disallowance*, the state of the pressure sensor, and the robot's motivational state against each other. An example of the result of this reconstruction is

displayed in figure 5.9. On the top blue line we can see the timing of four instances of *prohibition*$^{+}$. During the second of the displayed utterances the participant starts to restrain the robot's arm movement (1st peak middle red line), which lead to a transition of the robot's motivational state from positive to negative (bottom green line).

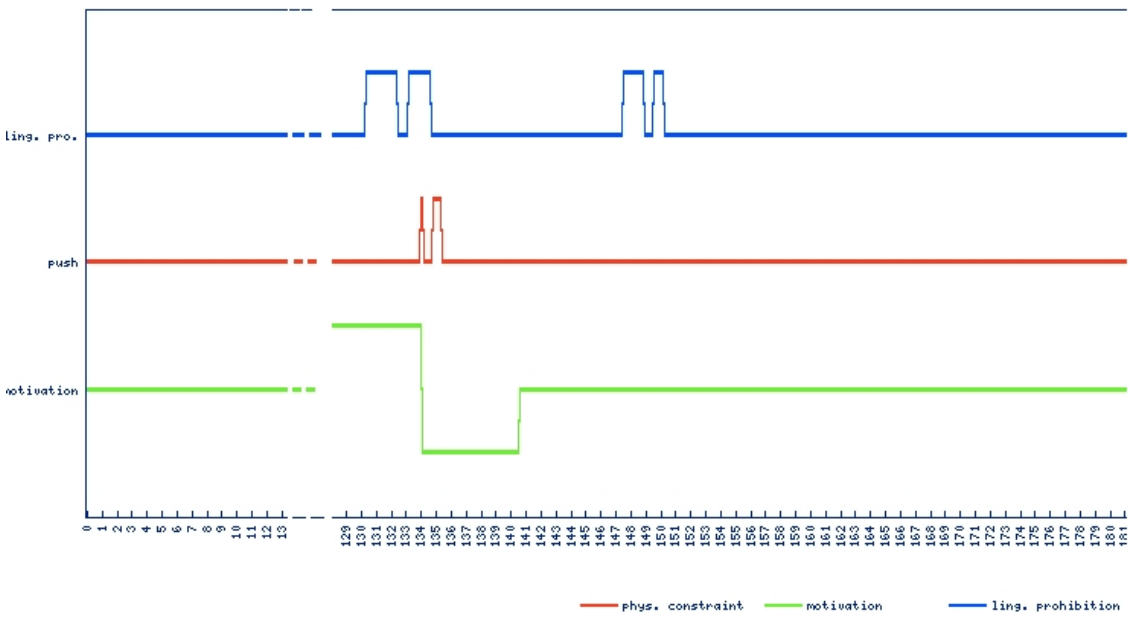


Figure 5.9: ***Excerpt of reconstructed temporal profile of human-robot interaction****: the given excerpt, taken from the reconstructed profile of P14's 3rd session, displays the temporal relation between* prohibitive *utterances and utterances of* disallowance *(top blue line), the robot's sensing of pressure being applied to its arm (middle red line), and the robot's internal motivation (bottom green line).*

During the manual analysis of these temporal profiles we observed the temporal relations depicted in figure 5.10. Additionally we observed the relations *between pushes*, and *during several pushes*. *Between pushes* can be decomposed into:

$$\text{between\_pushes} \Leftrightarrow \text{after\_push}(i) \wedge \text{before\_push}(j) \text{ with } j > i$$

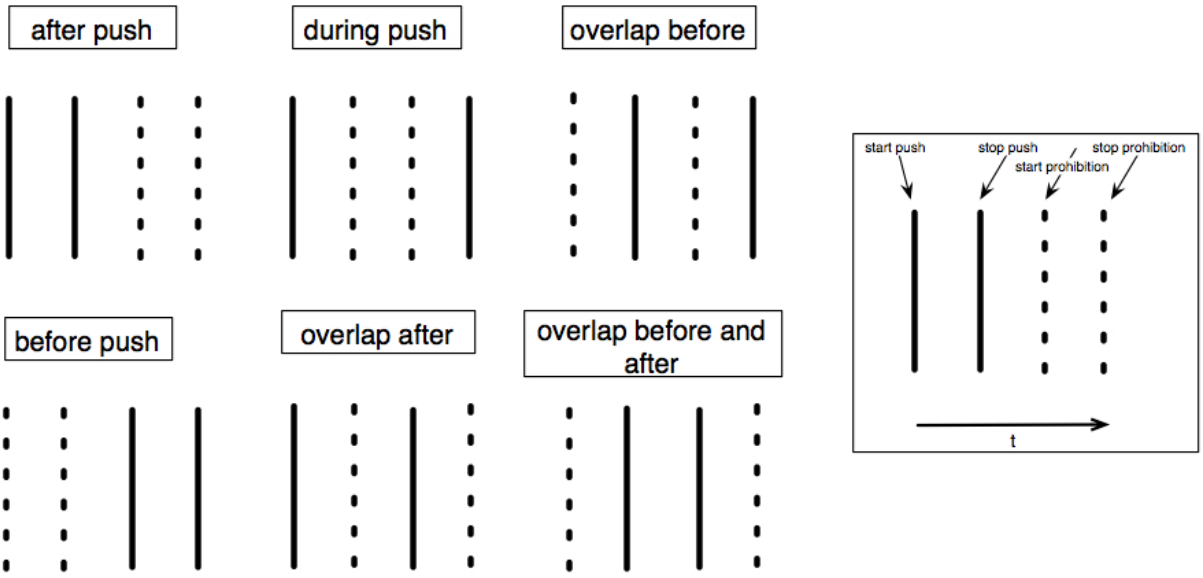


Figure 5.10: ***Basic temporal relations between corporal constraints and prohibition$^+$:*** *The depicted temporal relations between* prohibition$^+$ *and corporal constraints ("push") were observed within the prohibition scenario. Additionally two complex relations were observed which can be decomposed in the depicted ones (see text).*

and where the indices in brackets refer to the temporal order of *push* actions. *During several pushes* can be decomposed into:

$$\text{overlap\_after}(i) \wedge \text{overlap\_before}(j) \text{ with } j > i.$$

Important in this context are temporal constraints that apply to the *prohibition$^+$-push* relations. Assuming that a participant did execute at least two *pushes* during a particular session and assuming further that the same participant produced at least one *prohibition$^+$* in between the two *pushes*, this *prohibition$^+$* will be naturally "between pushes", no matter how large the two gaps between the utterance and the two *pushes* are. We thus have to impose temporal constraints on the gap between corporal and linguistic action in order to render these relations meaningful. By looking at the various temporal profiles and

revisiting the video recordings we determined a temporal limit of 4 seconds as maximum gap size. If the gap between a *push* and a *prohibition*$^+$ (and vice versa) exceeds this limit the two actions are not considered to stand in any of our specified temporal relations. As a last addition we added a *no push* (non-)relation to our set of temporal relations in order to account for cases where participants produced *prohibitions*$^+$ without touching the robot at all.

5.5.2 Evaluation of temporal relations

After establishing which temporal relations prevailed between *push* actions on the one hand and *prohibitions*$^+$ on the other, a script was written to automatically extract and count these relations based on the files resulting from the pragmatic coding and the robot's log files of the first three sessions of our 10 participants within the prohibition experiment[52]. The relation counts based on the 30 file tuples are displayed in table 5.54.

As becomes evident from this table, our participants violated the simultaneity constraint quite dramatically. If we rank the totals displayed in the bottom row by frequency, the first two ranks are distributed as follows: for 6 participants *no push* was the most frequent (non-)relation, followed by 3 participants which most often uttered *prohibitions* and *disallowances* while restraining the robot (*during push*). On place 3 of the top-rank is *before push* for 2 participants.

[52]Remember that the prohibition task was only present during the first three sessions of this experiment.

Table 5.54: ***Count of temporal relationships between physical constraints and prohibitive utterances****. Given are the counts of observed temporal relationships. Both* prohibitions *as well as* disallowances *were taken into consideration in the given count. Counts are given for all participants and sessions in the prohibition scenario in which participants were told to physically restrain the robot in case of it approaching a forbidden object. Furthermore a total count per participants is given in the last section of the table. A missing relationship type in a session indicates that all counts were 0. Temporal relationships of the listed types set in bold are very likely to be detrimental for an association of the salient word with negative affect. Relationships of a type set in italic are less likely to be detrimental for said association depending on the length of the gap between push(es) and utterance and the hypothesized duration of the motivational state triggered by physical restraint.*

		P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
s1	**no_push**	0	0	15	14	1	4	6	0	1	4
	before_push	1	0	0	0	0	1	3	2	5	0
	overlap_before	1	0	0	0	1	0	0	0	1	0
	overlap_before_and_after	1	0	0	0	2	0	0	0	0	0
	after_push	0	0	1	0	1	0	1	2	3	0
	overlap_after	1	0	1	0	1	1	0	0	0	0
	between_pushes	0	0	0	0	1	0	0	2	1	0
	during_push	4	0	1	0	2	0	1	4	3	0
s2	**no_push**	0	3	5	10	0	0	0	3	30	2
	before_push	3	4	1	0	0	2	0	3	0	0
	overlap_before	2	0	1	0	1	1	0	1	0	0
	overlap_before_and_after	0	0	1	0	0	0	0	1	0	0
	after_push	0	1	0	0	0	0	0	1	0	0
	overlap_after	1	0	0	0	0	0	0	0	0	0
	between_pushes	0	0	0	0	0	0	0	1	0	0
	during_several_pushes	0	0	1	0	0	0	0	2	0	0
	during_push	5	0	2	0	4	2	0	2	0	0
s3	**no_push**	0	2	1	3	0	1	0	3	30	5
	before_push	0	3	2	0	0	1	3	2	0	0
	overlap_before	1	1	6	3	3	0	1	2	0	0
	overlap_before_and_after	0	1	0	0	0	1	0	1	0	0
	after_push	0	1	0	1	1	1	0	0	0	0
	overlap_after	0	0	2	1	1	0	0	1	0	0
	between_pushes	0	0	5	1	4	0	0	0	0	0
	during_several_pushes	0	0	0	0	0	0	0	1	0	0
	during_push	2	0	8	1	8	0	2	5	0	0
total	**no_push**	0	5	21	27	1	5	6	6	61	11
	before_push	4	7	3	0	0	4	6	7	5	0
	overlap_before	4	1	7	3	5	1	1	3	1	0
	overlap_before_and_after	1	1	1	0	2	1	0	2	0	0
	after_push	0	2	1	1	2	1	1	3	3	0
	overlap_after	2	0	3	1	2	1	0	1	0	0
	between_pushes	0	0	5	1	5	0	0	3	1	0
	during_several_pushes	0	0	1	0	0	0	0	3	0	0
	during_push	11	0	11	1	14	2	3	11	3	0

The second rank is distributed as follows: For 4 participants *before push* is the second-most frequent relation, followed by *overlap before* (3 participants) and *during push* (2 participants).

How temporal relations of prohibitive action impact on the acquisition of negation

In order to understand why this frequent constraint violation is detrimental for the acquisition of negation, we will discuss this with some examples after a short recapitulation of architectural design targets and features which are important to understand the acquisition dynamics.

Let us assume *no* as default negative word for the current exposition and let us further assume that *no* is the salient word of the $prohibitions^{+}$ produced by the participant.

The hypothetical assumption that drove the design of both our architecture as well as our experiments is that the robot's internal negative motivational state can be associated with negative words similar to the association of object labels with perceptual features in other symbol-grounding architectures. In memory-based learners such an association comes to be merely by virtue of having a majority of exemplars in the memory where this association is established. For our purposes this means that, all other things, i.e. sensorimotor-states, being equal, such an association is established as soon as the majority of *no*'s are attached to sensorimotor-motivational data with a negative motivational entry. This means for the following exposition that any temporal relation which leads to a *no* with negative motivational value attached being added to the lexicon is beneficial for our learning target (**good**). By contrast any temporal relation which leads to a *no* with positive motivational value being added is detrimental to this purpose (**bad**).

Our version of symbol grounding is implemented such that the salient word is associated

with all variants of the sensorimotor-motivational (*smm*) vector that co-occurred during the time when the utterance was produced. During short time frames of a few seconds, the typical length of an utterance, in most cases nothing in this vector changes: The robot's behaviour does not change, the presented object stays the same, **and**, importantly, the object is recognized by the object detection to be the same. If this is the case while participants produce an utterance, the outcome will be one additional exemplar or lexical entry that is added to the robot's embodied lexicon. Yet, if one *smm* change occurs during this production, two lexical entries for the same word will be created, one for each variant of the *smm* vector. Changes in the *smm* data are caused through changes of the robot's behaviour, which for their part are caused by either timeouts or changes in the object recognition. Also changes in the object id itself are forms of sensorimotor changes and so is the change of the robot's motivational state. We will in the following assume that the object recognition works perfectly[53].

The robot's behaviour was implemented such that it would only grasp for objects that it likes, i.e. objects that cause its motivational state to be positive. Subsequently, in the vast majority of cases within the prohibitive setup, the robot's motivational state will be positive before the participant restrains its arm movement (push action)[54]. Restraints of the robot's agency lead immediately to Deechee becoming 'grumpy', i.e. a negative motivational state.

For *no push* relations the following happens: *No* is uttered while the robot is and continues to be in a 'positive mood', for its agency is not impeded. Instead of restraining the robot's arm, as they were taught to do, participants often just held the object out of

[53]Nothing could be further from the truth ...

[54]We say 'in the majority of cases' here, because it is possible due to a suboptimal object detection that a presented object is misrecognised as a non-desirable one just before the application of restraint. This would normally lead to the robot stopping to grasp. Yet if the misrecognition 'kicks in' just before the participant starts to push the robot's arm and therefore before the robot retracts its arm, its motivational state could be other than positive.

the robot's reach, which has no impact on its motivational state. Such interaction will lead to at least one exemplar of *no* in the robot's lexicon which is associated to a *smm* vector which has a positive motivational entry. This is **bad**.

In contrast, Deechee will already be 'in a negative mood' in case of participants starting to restrain its arm before uttering a $prohibition^{+}$ (*during push*). In this case one embodied word will enter the lexicon: *no* associated with a *smm* vector containing a negative motivational value. This is how we imagined the interaction to proceed from an architectural standpoint motivated by assumptions of simultaneity in ostensive theories of meaning. This is **good**.

In case of a participant starting to produce an utterance and subsequently constraining the robot's arm movement during that production (*overlap before*), two lexical entries will be created: a *no*, associated with a *smm* vector with a positive motivational entry, and additionally a *no*, which is associated with an otherwise identical *smm* vector but with a negative motivational entry. This is **in-between**.

If the onset of utterance production happens during a *push* but extends to after the end of the push (*overlap after*), the result will be one additional *no* in Deechee's lexicon associated to a *smm* vector with negative motivational entry *as long as the utterance is not overly long*. The robot's motivational system is implemented such, that its motivational state has a certain time lag. The only exception to this rule are restrictions of Deechee's freedom of movement which will make it grumpy immediately. This is most probably **good**.

These examples should give the reader an idea, how these different temporal relations affect the robot's lexicon.

Quantification of the robot's motivational state during prohibition$^+$

In order to verify the just given explanations and quantify more precisely which motivational states prevailed within the robot while participants engaged in *prohibition*$^+$ we counted the motivational states for each temporal relation and participant. The results are displayed in table 5.55. The tabulation largely corroborates our explanations and further shows that not many unpredictable recognition or other noise distorted the data. If we add up for each participant the positive and negative motivational states prevailing during the production of *prohibitions*$^+$ it becomes clear that our expectations did not materialise. In the interaction of 7 out of 10 participants the robot was more often in a positive motivational state when being prohibited as compared to it being in a negative state. Only for three participants, *P13*, *P17*, and *P20*, the opposite holds true, and for *P20* the margin of the robot being in a negative over being in a positive motivational state is rather small.

We suspect this lacking simultaneity between the robot's negative motivational state and the participant's production of *prohibition*$^+$ to be the reason, probably the major reason, for the acquisition of negation to be less successful within the prohibition experiment as compared to the rejection experiment.

Table 5.55: ***Motivational states during utterances of prohibition and disallowance***. *Given are the counts of the robot's motivational states for each temporal relationship between prohibition/disallowance and physical restraint. The counts are listed per participant within the prohibition experiment (see table 5.54 for the frequencies of these relationships). The counts are accumulated over the first three sessions, i.e. the sessions in which physical restraint could possibly occur. Note that one occurrence of such a temporal relationship can yield more than 1 to the count (see section 5.5.2 for details). The entries for* P13 *for* overlap_before_and_after *is so big due to a glitch in the motivational and/or behavioural system. Symbols used: -: negative motivation, +: positive motivation,* O*: neutral motivation*

	P13			P14			P15			P16			P17		
	-	O	+	-	O	+	-	O	+	-	O	+	-	O	+
no_push	0	0	0	0	0	5	4	4	17	0	1	14	0	0	1
before_push	0	1	4	0	0	7	0	0	3	0	0	0	0	0	0
overlap_before	4	0	4	1	0	1	6	0	6	3	0	3	5	0	3
overlap_before_and_after	11	12	0	1	0	1	1	0	1	0	0	0	2	0	2
after_push	0	0	0	2	0	1	1	0	0	0	0	1	2	0	0
overlap_after	2	0	0	0	0	0	3	1	0	1	0	0	2	0	0
between_pushes	0	0	0	0	0	0	2	0	3	1	1	1	4	0	2
during_several_pushes	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
during_push	11	0	0	0	0	0	11	0	0	1	0	0	14	0	0

	P18			P19			P20			P21			P22		
	-	O	+	-	O	+	-	O	+	-	O	+	-	O	+
no_push	0	0	5	0	2	4	2	4	1	3	14	16	0	1	7
before_push	0	0	4	0	0	6	0	0	7	0	0	5	0	0	0
overlap_before	1	0	1	1	0	1	3	0	3	1	0	1	0	0	0
overlap_before_and_after	1	0	1	0	0	0	2	0	2	0	0	0	0	0	0
after_push	1	0	0	1	0	1	3	0	0	3	0	0	0	0	0
overlap_after	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
between_pushes	0	0	0	0	0	0	2	0	1	0	0	1	0	0	0
during_several_pushes	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
during_push	2	0	0	3	0	0	11	0	0	3	0	0	0	0	0

If we further take our participants' behaviour in terms of the temporal relationship between corporal and linguistic action as indicative of how parents interact with their children when prohibiting them from doing something, these observations have far reaching implications for the fundamental design of symbol grounding systems. We will discuss these implications later in the discussion chapter 6.

Quantification of the robot's motivational state during other frequently produced forms of negation

After realising how often participants violated the simultaneity constraint in the context of *prohibition*$^{+}$, we are forced to ask ourselves if this constraint is heeded by our participants when engaging in any of the other frequent negation types. To this purpose we performed the same reconstructive temporal alignment between the robot's motivational states and the second group of major sources of negation words: *negative intent interpretations* and *negative motivational questions*.

The tables 5.56 and 5.57 display the numbers and kinds of motivational states of the robot during all of the participants' productions of *negative intent interpretations* and *negative motivational questions*. A manual inspection of these tables shows that in only 7 out of 100 sessions the number of positive motivational states outnumbered the negative ones.

Table 5.56: ***Motivational states during negative intent interpretations and neg. mot. questions within rejection experiment.*** *Given are the counts/number of associations of the robot's motivational states per stated utterances types. These frequencies are listed per session and accumulated across sessions. Symbols used: #: number of occurrences of the stated utterance type, -: frequency of negative motivational state, +: frequency of positive motivational state,* O*: frequency of neutral motivational state.*

		P01				P04				P05				P06				P07			
		#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+
s1	neg. int. int.	1	1	0	0	6	6	0	0	17	11	0	7	5	4	1	1	12	10	2	0
	neg. mot. question	0	0	0	0	3	3	0	0	6	1	3	2	3	2	0	1	13	5	7	2
s2	neg. int. int.	1	1	0	0	6	6	1	0	7	2	3	2	8	4	3	1	16	8	8	0
	neg. mot. question	0	0	0	0	7	7	0	0	11	7	4	1	1	1	1	0	7	1	5	2
s3	neg. int. int.	0	0	0	0	3	3	0	0	6	4	3	0	12	12	2	0	7	5	2	0
	neg. mot. question	0	0	0	0	5	3	2	0	2	0	2	1	2	2	0	0	18	2	9	9
s4	neg. int. int.	0	0	0	0	0	0	0	0	3	2	0	1	13	12	3	1	7	6	0	2
	neg. mot. question	0	0	0	0	1	0	1	0	5	4	1	1	3	2	1	0	10	0	6	5
s5	neg. int. int.	0	0	0	0	8	7	0	1	3	1	2	0	11	8	5	0	7	6	2	0
	neg. mot. question	0	0	0	0	8	8	0	0	6	4	3	1	1	0	1	0	24	5	14	7
total	neg. int. int.	2	2	0	0	23	22	1	1	36	20	8	10	49	40	14	3	49	35	14	2
	neg. mot. question	0	0	0	0	24	21	3	0	30	16	13	6	10	7	3	1	72	13	41	25

		P08				P09				P10				P11				P12			
		#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+
s1	neg. int. int.	6	3	3	0	6	6	1	1	0	0	0	0	9	6	4	1	7	4	2	3
	neg. mot. question	10	1	9	1	4	3	2	0	0	0	0	0	3	2	0	1	2	0	2	0
s2	neg. int. int.	3	2	1	0	6	4	4	0	0	0	0	0	4	3	1	0	2	1	1	0
	neg. mot. question	7	1	6	0	5	2	1	2	2	2	0	0	0	0	0	0	2	1	1	0
s3	neg. int. int.	1	1	0	0	6	4	1	1	1	1	1	1	6	6	0	0	0	0	0	0
	neg. mot. question	10	6	1	3	7	3	6	1	1	1	0	0	2	2	1	0	0	0	0	0
s4	neg. int. int.	4	3	1	0	5	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0
	neg. mot. question	10	6	4	0	0	0	0	0	3	3	3	2	2	1	0	1	0	0	0	0
s5	neg. int. int.	4	3	2	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	neg. mot. question	6	3	2	1	4	1	1	3	3	4	2	0	0	0	0	0	0	0	0	0
total	neg. int. int.	18	12	7	0	25	18	11	2	1	1	1	1	19	15	5	1	9	5	3	3
	neg. mot. question	43	17	22	5	20	9	10	6	9	10	5	2	7	5	1	2	4	1	3	0

Table 5.57: ***Motivational states during negative intent interpretations and neg. mot. questions within prohibition experiment***. *Given are the counts/number of associations of the robot's motivational states per stated utterances types. These frequencies are listed per session and accumulated across sessions. Symbols used: #: number of occurrences of the stated utterance type, -: frequency of negative motivational state, +: frequency of positive motivational state,* O*: frequency of neutral motivational state.*

		P13				P14				P15				P16				P17			
		#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+
s1	neg. int. int.	6	5	3	1	0	0	0	0	15	9	7	0	7	2	4	1	4	3	1	0
	neg. mot. question	4	2	1	1	0	0	0	0	11	5	6	1	3	2	2	0	1	1	0	0
s2	neg. int. int.	6	6	0	0	0	0	0	0	4	4	0	0	4	3	2	0	4	4	1	0
	neg. mot. question	1	0	0	1	0	0	0	0	8	2	4	3	3	1	2	0	1	1	0	0
s3	neg. int. int.	0	0	0	0	0	0	0	0	4	3	1	0	2	1	1	0	1	1	0	0
	neg. mot. question	1	1	0	0	0	0	0	0	6	4	3	1	1	0	1	0	1	0	1	0
s4	neg. int. int.	1	1	0	0	0	0	0	0	8	5	4	1	5	2	4	1	1	1	0	0
	neg. mot. question	4	4	0	0	0	0	0	0	9	4	4	2	4	1	3	1	0	0	0	0
s5	neg. int. int.	2	2	0	0	0	0	0	0	7	5	1	1	13	4	9	0	3	2	1	0
	neg. mot. question	2	2	0	0	0	0	0	0	18	9	8	2	3	3	0	0	4	2	2	0
total	neg. int. int.	15	14	3	1	0	0	0	0	38	26	13	2	31	12	20	2	13	11	3	0
	neg. mot. question	12	9	1	2	0	0	0	0	52	24	25	9	14	7	8	1	7	4	3	0

		P18				P19				P20				P21				P22			
		#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+	#	-	O	+
s1	neg. int. int.	15	9	4	3	4	3	1	0	8	4	5	0	1	1	0	0	2	0	1	1
	neg. mot. question	8	7	2	1	4	4	1	0	2	1	1	1	3	1	3	0	0	0	0	0
s2	neg. int. int.	4	4	0	0	4	4	0	0	1	1	0	0	0	0	0	0	0	0	0	0
	neg. mot. question	2	1	1	0	3	3	1	0	10	7	4	0	3	3	0	0	0	0	0	0
s3	neg. int. int.	6	4	3	0	5	5	0	0	5	4	2	1	0	0	0	0	0	0	0	0
	neg. mot. question	2	0	2	0	0	0	0	0	7	5	3	0	3	2	0	1	0	0	0	0
s4	neg. int. int.	2	2	0	0	1	1	0	0	6	4	2	0	0	0	0	0	0	0	0	0
	neg. mot. question	1	1	0	0	3	3	0	0	6	5	2	1	3	2	1	0	0	0	0	0
s5	neg. int. int.	3	2	0	1	4	4	1	0	2	2	1	0	3	3	0	0	0	0	0	0
	neg. mot. question	1	1	0	0	2	2	0	0	13	10	3	1	8	7	1	0	0	0	0	0
total	neg. int. int.	30	21	7	4	18	17	2	0	22	15	10	1	4	4	0	0	2	0	1	1
	neg. mot. question	14	10	5	1	12	12	2	0	38	28	13	3	20	15	5	1	0	0	0	0

Statistical comparison of prevailing motivational states during the production of negation In this paragraph we seek to verify the impression that *negative intent interpretations* and *negative motivational questions* were indeed expressed considerably more often while the robot was in a negative motivational state as was the case during the production of *prohibitions* or *disagreements*. To this purpose we compare the relative frequencies of all three motivational states of the robot as measured for each participant and type accumulated over all sessions. In other words, we computed for each participant and type the relative frequencies (percentages) of each motivational state (cf. table B.30). In order to statistically compare the relative frequencies of motivational states between the three negation types, the relative frequencies pertaining to a particular motivational type, *negative*, *neutral*, *positive*, and pertaining to a particular negation type for all participants were pooled into one set. The ensuing nine sets, one set per motivational state and negation type, were subsequently compared separately for each motivational state, resulting in 3 separate ANOVAs[55].

In order to illustrate the comparison, an example shall be in order: For finding out if it is indeed the case that participants produced *negative intent interpretations* significantly more often while the robot was in a negative motivational state as compared to them having produced *prohibitions* while the robot was in a negative motivational state, we compared the relative frequencies of negative motivational states of all participants between the three types applying an ANOVA to these sets. I.e. we compared the relative frequencies stated in the 1st column (not counting the column of participant ids) of table B.30, with the entries of the 5th and 9th column.

The ANOVAs were performed under two different conditions. In *condition 1* none of

[55] As the relative frequencies of the motivational states for one particular type are mutually dependent, one ANOVA would have presumably sufficed. But due to our shallow statistical knowledge, we wanted to be sure to cover all eventualities.

Table 5.58: ***Statistical comparison of relative frequencies of motivational states between negation types.*** *Displayed are the mean and standard deviation of the relative frequencies of motivational states in percent grouped by negation type. Furthermore the F-values resulting from the ANOVA comparing the relative frequencies across types is presented.* Condition 1: *no exclusion of data,* condition 2*: exclusion of relative frequencies based on less than 10 utterances. Abbrev.: -: neg. mot. states,* O*: neut. mot. states, +: pos. mot states.* $\star : p < 0.0001$, $\dagger : p < 0.01$, $\ddagger : p < 0.02$

	Condition 1				Condition 2			
	NII	NMQ	P	*F*	NII	NMQ	P	*F*
-	64.34 (24.65)	54.99 (19.94)	36.70 (23.38)	$4.84^{\ddagger}$	65.94 (15.2)	56.44 (20.3)	40.78 (20.7)	$5.17^{\ddagger}$
O	25.45 (15.5)	34.79 (16.8)	10.35 (10.5)	$8.25^{\dagger}$	26.68 (12.9)	32.04 (14.2)	10.11 (11.1)	$8.05^{\ddagger}$
+	10.22 (13.8)	10.22 (8.9)	52.96 (24.2)	$30.78^{\star}$	7.39 (8.46)	11.52 (8.9)	49.12 (22.2)	$32.13^{\star}$

the frequencies were excluded from the data sets apart for those cases where the respective participant never produced the respective negation type such as participant *P14* for *negative intent interpretations* (cf. table B.29). In *condition 2* we excluded all relative frequencies which are based on less than 10 utterances.

The results of the statistical comparisons are presented in table 5.58[56].

As can be seen there, all differences in the mean relative frequencies are statistically significant. Whereas under both conditions participants on average produced *negative intent interpretations* (*NIIs*) and *negative motivational questions* (*NMQs*) about 10% of the time when the robot is in a positive motivational state, they produced *prohibitions*$^{+}$ while the robot is in that state at least 49% of the time. In contrast, most often participants produce *NIIs* with the robot being in a negative motivational state (∼64% vs. ∼66% under crit. 1 and 2 resp.[57]), i.e. most often their negative linguistic productions of this

[56]The given table is based on the data of all participants. We performed the same test for participants of the prohibitive experiment only. The resulting means were very similar and the differences of the relative frequencies between types were equally significant. We further tested for differences in the means of relative motivational frequencies of the same type across experiments, i.e. differences between *NIIs* produced within the rejection experiment and those produced within the prohibition experiment, using t-tests. None of the differences that were found were statistically significant. The same holds true for a cross-experiment comparison of relative motivational frequencies of *NMQs*.

[57]We will keep this order of percentages, the mean percentage under criterion 1 first followed by the

type are performed simultaneously with the robot's physical display. The same, although somewhat weaker, holds true for *NMQs* (∼55% vs. ∼56%), but not for the production of *prohibition*$^+$ (∼37% vs. ∼41%).

Another noteworthy observation is the statistically significant higher levels of productions of both *NIIs* (∼25% vs. ∼27%) and *NMQs* (∼35% vs. ∼32%) while the robot is in a neutral state compared to the production of *prohibitions*$^+$, which rarely takes place during a neutral motivational state (∼10% under both criteria). The explanation for this difference is straightforward: as explained earlier, the robot's typical motivations in the context of *prohibition*$^+$ are either positive or negative because participants typically engage in this type as soon as the robot approaches a forbidden object and the robot only does this when being in a "positive mood". As soon as participants push the robot's arm, its motivational state flips immediately to negative, skipping the neutral state. We assume the on average 10% of utterances, which were produced while the robot was in a neutral state, to be mainly instances of *disallowance*. The latter were also produced in less strict temporal alignment with prohibitive situations. The comparatively high percentage of instances of *NMQ* that were uttered while the robot was in a neutral state might be explained by some of the participants' interpretation of this state as Deechee being undecided. In such cases *negative motivational questions* might be meant as proper questions, as the robot indeed does not indicate his inclination towards an object in any direction.

mean percentage under criterion 2, in the following without saying so explicitly.

5.6 Summary

Utterance level

We started the analysis of our participants' speech on the utterance level (section 5.1). Three measures in particular were considered: the mean length of utterance (MLU), the number of distinct words ($\#dw$), and the number of utterances per minute (u/min).

In-group analysis First we determined whether any vertical changes could be observed, i.e. potential adaptations of participants in their way of speaking to the robot across the five experimental sessions. For the rejection experiment we found such adaptations for both MLU as well as $\#dw$. In the case of MLU these changes were only statistically significant when we compared the very first with the very last session. 3 out of 10 participants changed their $MLUs$ in a statistically significant manner and 2 further participants showed tendencies to do so. These changes were such that those participants with initially high $MLUs$ shortened their utterances across the sessions and, conversely, those participants with low $MLUs$ increased their utterance lengths. For the *number of distinct words* we observed a statistically significant drop by approximately 25, from approx. 117 to approx. 92, between the first two sessions after we excluded 2 participants as outliers. The two outliers on the other hand, both of which started out with extremely low $\#dw$ (6 and 21) increased this value between the first two sessions.

Within the prohibition experiment we observed a significant drop from approx. 4.5 to 2.3 words in the number of *distinct negative words* between the third and fourth session, i.e. at the transition point when the prohibitive task (the treatment) ceased to be.

Cross-group analysis When comparing the two negation experiments in terms of the abovementioned measures, we found statistically conspicuous differences in two of

them.

Participants of the rejection experiment appeared to generally use slightly shorter utterances (MLU: 3.13) than their peers from the prohibition experiment (MLU: 3.4). Yet this difference shows only a statistical tendency in the fourth session.

Considerably more substantial in statistical terms is the difference in the *utterances per minute* between the two experiments. Albeit participants of the rejection and prohibition experiment start out on average on approximately the same level (23.34 vs. 22.66 *u/min* in session 1), the 'rejective' participants talk less and less from session to session, whereas their 'prohibitive' peers do the opposite and appear to become more and more talkative as the experiment proceeds. The difference in talkativeness thus increases from session to session. Both groups finish the experiment with 21.58 *u/min* and 30.08 *u/min* respectively, rendering this difference statistically significant.

There appears to be a similar trend during the first three sessions in the subset of negative utterances only, yet the difference in the number of negative utterances does not reach a significant level. Upon cessation of the prohibitive task in session 4, participants decrease the number of negative utterances per minute such that the difference in this measure between the two groups becomes increasingly smaller.

Comparison to speech from Saunders' experiment When looking at the number of *utterances per minute*, we find a statistically increasing trend, as the experiment progresses, that participants of Saunders' experiment are located approximately in between the participants of the rejection and prohibition experiment.

A stark, and statistically significant difference from the very start is observable in the number of *negative utterances per minute* (*#nu/min*). On average participants from the rejection experiment produce more than 4 times as many negative utterances (3.2

#nu/min) compared to the participants of Saunders' experiment (0.74 *#nu/min*). This difference is even more pronounced when the latter group is compared to participants of the prohibition experiment (3.63 *#nu/min*). Here we see a first sign of the impact of the robot's display of emotion or volition upon the way people speak to the robot.

Word level

On the word level we established a ranking of word frequencies for each of the two word corpora that result from conceiving of the recorded speech of all participants from each experiment as one corpus - *PC* and *RC* for the rejective and prohibitive group respectively. Indicators of personal involvement such as the vocative are on high ranks in both the *PC* and *RC*. Moreover the vocative has a saliency rate well above average and would rank amongst the 10 most frequent salient words in the *RC*, had we transcribed it with one phonetic variant. We furthermore find very high frequencies of words such as *like* and *no* that are probably indicative of participants ascribing intentionality to the robot. In the *RC* *no* takes the 6th rank, thereby covering 2.39% of all words in the corpus compared to a coverage of 0.56% in the British National Corpus of spoken British English (*BNCS*). When considering only the subset of prosodically salient words (*RCS*), *no* moves up to the second rank and covers 4.64% of latter corpus. Similar findings for *no* apply to the *PC(S)*.

When looking at the saliency rates of different word groups we found for the *RC(S)* that emotion words such as *sad*, *happy*, or *smile*, have the second-highest saliency rate (71.28%) after the group of object labels (73.81%). This indicates that participants emphasised these words prosodically, which in turn indicates an inclination of participants to ascribe these motivational states to the robot. Within the *PCS* this group ranks on place 5 and has a saliency rate of 55%.

In terms of negation words, we found *no*, *don't*, and *not* to be the major players within

the *RC*, with *no* being both the top-ranking and the most salient word of this group. In the *PC(S)* this group is joined by *can't*, ranking slightly below *not* in the *PC* but overtaking the latter in terms of rank in the subset of salient words due to its comparatively high saliency rate (39.8%).

When comparing these findings to the corpus derived from speech in Saunders' experiment (*SC*), we found that participants there used *you* considerably less as compared to both the *RC* and *PC* (10th rank in *SC*, 1st rank in both *RC* and *PC*). Equally their usages of *like* (approx. −75%) and the vocative (approx. −50%) are considerably lower, which indicates a lower degree of involvement with the robot.

In terms of negation words the difference between the two negation corpora and Saunders' corpus is most dramatic. What are highly frequent words within both *RC* and *PC*, *no*, *not*, *don't*, and *can't* (*PC* only), are hardly existent within the *SC*. Moreover, the few occurrences of *no* that we find within the *SC* are considerably less salient (35.71% (*SC*) vs. 58.03% (*RC*), 59% (*PC*)). The lower production rate of negative utterances in Saunders' experiment on the utterance level is thus mirrored in far lower ranks of negation words on the word level. The lower salience rate of the already much fewer productions of negation words then means, that these words hardly ever make it into the robot's active vocabulary.

An interesting contrast in this context is the circumstance that the salience rate of emotion words (essentially *smile*) within the *SCS* (60%) is located in between its respective salience rate of the *RC* (71.28%) and *PC* (54.9%). *Smile* is in terms of its relative frequency produced more often within Saunders' experiment (coverage: 0.35%) as compared to the *RC*(coverage: 0.07%) and *PC*(coverage: 0.08%). *Happy* and *sad* on the other hand, very frequently occurring emotion words within the negation corpora (*happy*: 38 (*RC*), 22 (*PC*), *sad*: 35 (*RC*), 8 (*PC*)), were not or hardly produced within Saunders' experiment (*happy*: 1, *sad*: 0).

Pragmatic level

First, we presented the two negation taxonomies whose construction was guided by Pea's taxonomy of early child language negation but which are otherwise based on the forms of negation which we encountered in the conversations between participants and robot.

Evaluation of the taxonomies Subsequently we presented the evaluation of the taxonomies, which, for the lack of a pragmatic gold standard, is based on the agreement between two coders (intercoder agreement) that applied this taxonomy to a subset of the collected negative utterances. This evaluation yielded a good agreement for the human negation taxonomy ($\kappa = 0.74$), but only poor agreement for the taxonomy of negative robot utterances ($\kappa = 0.46$). Apart from letting the coders classify both the participants' as the robots' negative utterances for negation types, they also judged each of the robot's utterances as to whether they are felicitous or adequate in the respective situations. The agreement for these judgments turned out to be even lower ($\kappa = 0.41$) than the agreement with regard to the type. This value is at the lowest end of the boundary for moderate agreement of the most forgiving of the three scales considered.

In an attempt to improve the robot negation taxonomy we developed and applied two complementing methods. First we developed an automatic optimization algorithm for the taxonomy that is guided by the idea that a reduction of negation types by virtue of merging distinct types into hybrid types would necessarily lead to an improvement of agreement. With the second method, a structured interview of the 2nd coder, we determined the reasons for the coders' disagreement. The confusion matrix that resulted from the dual coding indicated that the major source of disagreement between coders pertained to the coders' choice between the types *truth-functional denial* and *rejection*. Also the automatic optimization suggested to fuse precisely these two types into a hybrid-type. As this fusion

would render the ensuing taxonomy virtually incomparable to established taxonomies for early child negation, we rejected this suggestion. Instead, we have to accept that the robot's physical and linguistic behaviour as it was executed during the experiments, is not sufficient for external coders to judge the robot's negative utterances sufficiently well according to type or adequacy.

Rather than artificially boosting the agreement by reducing the taxonomy in the suggested, and what we consider nonsensical, way, we decided instead to focus on the reasons why a higher agreement might not be reached with the given architecture, assuming the same taxonomy. The major reasons identified for the coders' disagreement was, firstly, a lack of turn-taking skills.

In particular the fact that the robot is deaf in real-time while interacting with participants and its lacking real-time sensitivity for other social signals such as gaze or prosody, renders it incapable to engage in conversational moves in a timely manner. Yet an important distinguishing criterion within our taxonomy is the linguistic and behavioural adjacency of utterances to preceding conversational turns. Thus, as the robot produced his turns under ignorance of preceding human turns, and as the temporal placement of conversational turns is an important criterion to identify the type of a negative utterance, coders had considerable problems in making decisions with regards to type as well as felicity.

Secondly, and especially important for judgments on felicity, the two coders often disagreed with regards to the robot's intentions. This was typically caused through a suboptimal inverse-kinematics module, which, at times, prevented the robot from reaching out for an object, or which lead to it twisting its arm in humanly impossible ways. In these cases, the first coder and architect, knew about the actual internal state of the robot as indicated by its (reliable) facial expressions, its 'real intention' to grasp, whereas the second coder often identified this behaviour as unwillingness to grasp. One could express

this insight alternatively as: When it comes to negation, the success of what one does (the speech act), depends on what one was intending to do. Conversely: If two persons have different opinions with regards to the goal of an actor, they judge the success of his moves under different criteria and therefore potentially differently. Taken out of context this insight sounds astonishingly trivial, yet, as we will show here in very concrete ways, it has far-reaching implications for the design of symbol-grounding architectures.

The only positive aspect when looking at the coders' judgments of felicity is that both coders yielded close to the same numbers of negative utterances which they considered felicitous. Thus, there is no indication that the first coder, whose judgment was subsequently used to determine the success of acquisition within the respective experiments, judged the robot's performance in an unduly positive way.

Pragmatic analysis The application of the human robot taxonomy to the utterances produced by participants within the rejection experiment showed that the three negation types, that participants from this experiment engaged in most, were *negative intent interpretations* (*NII*), *negative motivational questions* (*NMQ*), and *truth-functional denial* (*TFD*). The finding, that most negative utterances were instances of the aforementioned two motivational types (*NII*, and *NMQ*) corroborates the judgment, that we made on the utterance level: the participants' production of negative words was indeed linked to the robot's motivation-related behaviour.

We furthermore found that negation words which were produced within instances of these two motivational types had considerably higher salience rates as compared to those negation words which were produced within instances of *truth-functional denial* (rej. exp.: *NII*: 48.6%, *NMQ*: 54.3%, *TFD*: 29.1%). The combination of these two facts, high overall production rate and high salience rate of negation words, means that the majority of

negation words, above all *no*, which entered the robot's lexicon originated from human utterances of the two motivational types.

Within the prohibition experiment the top rank of the most frequently produced negation type was taken over by linguistic *prohibition* while productions of the aforementioned types still remained highly frequent. *Prohibitions* (P) were found to have the highest salience rate with regard to negative utterances (pro. exp.: P: 60.5%, NII: 38.2%, NMQ: 41.8%, TFD: 31.7%). Moreover we could not find a single prohibitive utterance within our corpus, that did not contain a negation word. This contrasts with the fact that intent interpretations were not necessarily all negative: we did observe formulations such as "Oh, you look sad" which appeared to have been uttered by some participants in exactly those situations where other participants resorted to negative variants such as "No, don't like it"[58]. The latter observation in combination with the fact that negative words have a very high salience rate with *prohibitions* renders *prohibitions*, at least in numerical terms, the best source of negative words.

Linking word- and pragmatic level In order to determine more precisely, where the negation words that entered the robot's lexicon originated, we correlated the most frequent negation words with the most frequent negation types. This correlative analysis showed that within those sessions, where the prohibitive task was present, linguistic *prohibitions* were the primary source of *no*, followed by instances of the other three frequent types (P: 129 (*85*)[59], NII: 78 (*48*), NMQ: 64 (*57*), TFD: 74 (*29*)), due to the combination of high frequency and high salience rate ($\sim 56-69\%$). Albeit instances of *truth-functional denials* had within the rejection experiment the highest absolute frequency of *no's*, the

[58]Due to our exclusive focus on negation and negative utterances, we don't know how positive and negative variants numerically relate to each other. This further means that the positive variants did not enter our statistics.

[59]Format: absolute count (*salient count*)

salience rate of this word within instances of this type in general ($\sim 35-37\%$) is considerably lower than the salience rate of *no's* produced within instances of the two motivational types (*NII*: $63-64\%$, *NMQ*: $80-87\%$). Within the rejection experiment it is therefore the two motivational types that constitute the major source of *no's* which end up in the robot's lexicon despite the higher absolute frequency of this word within *truth-functional denials* (*NII*: 97 (*62*), *NMQ*: 127 (*95*), *TFD*: 137 (*44*)).

Sources of salience After having determined the varying degrees of saliencies of negation words, with *no* typically having the highest rates amongst all negation types, we attempted to isolate the reason for this salience. We determined that *no* as opposed to other negation words, is typically part of shorter utterances many of which are one-word utterances. This circumstance contributes or may even be the major factor for *no* to be prosodically salient under the given operationalisation of prosodic saliency.

Assessment of the robot's performance

The robot's performance within the two negation experiments was assessed by comparing the percentages of felicitous negative utterances of the last two sessions of both experiments. The felicity (or adequacy) of Deechee's utterances was significantly worse in the prohibition experiment ($\sim 28-33\%$) compared to the rejection experiment ($\sim 64-74\%$). This result did surprise us, especially considering the fact that linguistic *prohibitions* appeared to be an excellent source of negation.

Temporal relationship between corporal and linguistic prohibition

In order to explain the mismatch between our intuition about prohibitions as good sources of negation and the worse performance of the robot within the prohibition experiment, we focused on one constraint of the acquisition algorithm, which we termed simultaneity

constraint: the simple assumption that 'referent' and word occur roughly simultaneously. As our analysis showed, most participants in the prohibition experiment did indeed violate this constraint. 6 out of 10 participants most often acted against the instructions and did not restrain the robot's movement at all. Another frequently, and for our algorithm detrimental, observed temporal ordering between linguistic and physical prohibition, was the execution of the linguistic act before physical restraint was applied. Only 3 participants followed our instruction in the majority of cases. A subsequent analysis of the temporal alignment between the robot's motivational state and prohibitive utterances showed that the robot was on average more often in a positive ($49-53\%$) than in a negative motivational state ($37-41\%$) while participants uttered *prohibitions*.

By contrast we found that *negative intent interpretations* were produced in $64-66\%$ of the cases while the robot was in a negative motivational state as opposed to $7-10\%$ of the cases in which it was in a positive state. Similarly *negative motivational questions* were $55-56\%$ of the time produced while the robot was in a negative motivational state and only in $10-12\%$ of the time while it was in a positive state.

The fact that the majority of prohibitions are uttered while the robot is in a positive state fundamentally prevents any associative algorithm whose functioning depends on approximate temporal simultaneity of symbol and referent to establish the correct association. Yet before prematurely abandoning *prohibitions* as hypothetical early sources for the child's negation, we will pick up this point in the discussion chapter (chapter 6) as this might give us important clues about principal shortcomings of the current generation of symbol-grounding algorithms.

P12 ((holds out crescent box towards D))
(1.5)
D ((starts to frown))
(1.8)
D crescent
P12 yea::h crescent well done
P12 ((puts down crescent box quickly))
D no ((P12 picks up circle box))
(.9)
P12 <stop saying no> ((laughs))
—*Participant 12 and Deechee, session 5*

Chapter 6

Discussion

6.1 What did we learn?

6.1.1 Does the robot now really know how to say 'no'?

A very simple sounding question that one may pose to us at this point is whether the robot now knows how to negate. The answer to this question, lamentably, is far less simple. We have seen in the previous chapter that the felicity rates of the robot's negative utterances were between 64% and 74% in the case of the rejection experiment, but only between 28% and 33% in the prohibition experiment. Superficially one could now claim that in the former experiment the robot by and large has evidently learned how to negate, as far as an external oberver can tell, whereas it did not in the latter experiment.

The first issue here is that we have no naturalistic numbers to compare our results to. None of the authors of our negation taxonomies have provided us with felicity rates for the utterances of the children under observation. Even if they had given us felicity rates, we presumably would not know how reliable they would be, as these authors did not employ any second coders to evaluate the reliability of their judgements. This then means, that

we really don't know how good 30% or 70% are.

A plea towards the psycholinguists Thus, from our perspective, the first task for modern psycholinguists is to repeat these experiments and observations and employ modern methodology: multiple coders, and quantitative results with regards to negation type as well as to felicity. What is also needed is more detailed and quantitative characterisation of the source of error, in case of low intercoder agreement and/or low felicity rates.

At least one attempt has been made to evaluate the coder reliability for communicative functions of very young children. We would like to remind the reader that we explained our comparatively low intercoder reliabilty with the circumstance that the robot was 'deaf in real-time'. As it did not 'know' whether it was addressed or not, its utterances were often 'off' with regards to the timing. For this reason the coders often could not agree as to whether some utterance was 'really' an answer to a question, e.g. a case of truth-functional denial, or if it was to be taken as a non-adjacent negation type, e.g. motivation-dependent exclamation. This uncertainty subsequently impacts on decisions on felicity because the various negation types have differing felicity criteria. The same holds true for judgements on communicative intent, because communicative intent hinges, at least partially, on adjacency: a question 'has' or embodies a different intent than does a response.

Accordingly, if we are to undertake a comparison between robot and human child, we have to be fair and compare the robot with deaf children. Nicholas et al. (1999) performed a comparison of inter-coder reliabilities between coders that coded for communicative function and intentionality of deaf children of various ages and coders that coded for the same features of normally-hearing children. Their 'taxonomy' for communicative functions contained 10 different types such as *statement*, *directive*, *question*, *response*, etc., but also *no clear function*. Interestingly their coder agreement for intentionality for deaf children

at the age of 12 and 18 months was slightly above and slightly below 50% respectively, whereas it was slightly below 50% but nearly 75% for normally-hearing children of the same two age groups. Their coder agreement for the communicative function assignment for deaf children was first around 75% for 12 month old deaf children but then dropped to just over 50% for 18 month old deafs and appeared to stay that way until they were 54 months of age - the maximum age considered in the experiment. The coder agreement for the assignment of communicative function for normally-hearing children on the other hand started out just above the 50% mark for 12 month old children, but subsequently increased steadily towards the 75% mark at about 54 months of age. This is a very interesting result as the intercoder reliability, though measured here as relative agreement, is not far away from the one we measured for the robot's utterances.

According to Nicholas et al. (1999) their results indicated that "even for those acts which coders agreed were intentional, another 30-40% were ambiguous when coders attempted to discern communicative function. This degree of ambiguity is greater for hearing-impaired children, who presumably are less skilled communicators" (Nicholas et al. 1999, p. 128).

The good news then is, that our intercoder reliability for negative utterances is not that much worse as compared to the one yielded when coding for the speech act type of 12 to 18 month old deaf children. The bad news is that we have to be extremely careful with any judgement of a single author qua coder with regards to the frequency of communicative functions at early ages. Thus, had the authors of our negation taxonomies produced quantitative judgements about the type frequency and adequacy of their childrens' production of early negation, we would now have to be extremely critical about them.

This emphasises the need for a repetition of the observations with modern methods in order to verify the 'old' results, but also in order to quantify type frequency as well as adequacy or felicity. Only then we will get an impression of how good children are in

'bringing the message across', and only then we will be able to answer the question in the heading: did the robot really learn how to say 'no'?

6.1.2 Is hypothesis 2 now disqualified?

The short answer is: no. The reason for its non-disqualification due to the comparatively low felicity values in the prohibition experiment is that there might be another explanation, that does not apply to parent-child interactions. As we have seen in section 5.5, our participants often did not 'properly' push the robot's arm, i.e. they did not push it simultaneously to uttering the prohibition.

An innocent explanation An innocent reason for their frequent non-compliance with our instructions could be that they were either afraid or hesitant because they were not used to or felt uncomfortable touching the robot. In this case the 'misfit' of the timing between physical restraint and oral production of the prohibition may not apply to adult-child interactions. In other words, the synthetic model does not 'fit' its natural counterpart. In this case hypothesis 2 remains untouched and there are also no consequences for the symbol grounding system.

A less innocent explanation Far more problematic, not for hypothesis 2, but for our operationalisation of symbol grounding would be the following observation: what we have seen in the experiments is not dissimilar to what parents do with their children. What we have seen, was the frequent sole production of linguistic prohibition without the execution of the prescribed corporal restraint, or the temporally earlier production of linguistic prohibition before the application of corporal restraint. The high frequency of both of these 'events' lead to an 'incorrect' association between positive motivational state and negative word because the robot was still happy when the prohibition was uttered.

The reason for the robot still being happy was its deafness.

If parents often happen to shout before they touch, or if they happen to shout without touching at all, when uttering prohibitions, the timing becomes extremely important if a simple associative mechanism is assumed. It is easy to imagine, especially if we take speech act theory seriously, that shouting or the harsh pronunciation of an utterance is unpleasant for the child, i.e. causes negative emotions. In this case it could be that the phonetic decoding of the word takes the same amount of 'brain time' as it takes time for the 'emotional valence' of the utterance to unfold its impact. In this case a simple association mechanism still stands a chance. If the emotional impact is delayed as compared to the 'decoding' of the word, the word would be associated with a positive valence. If this is true *and* if hypothesis 2 holds, the learning algorithm is not one of simple association.

One could for example imagine that word as well as emotional valence are held for a few seconds in some memory, upon which the grounding qua association process operates. In this case the precise timing loses its importance. Many more cognitively more complex operations are imaginable in this context. The only way, it seems, to shed more light on the precise nature of the process, is with far more detailed observational data than is currently available. What is needed are detailed and long-term video recordings of parent-child interaction, not just in play settings, but rather in everyday settings where parental prohibition is likely to occur.

6.2 Summary of contributions

Within the work presented in this thesis we have tested two hypotheses pertaining to the origins of linguistic negation using a synthetic or constructive approach. The first step in approaches of this kind is to construct a robotic architecture that, at least to

a certain degree, mirrors and implements crucial aspects of the research question in the target field. In our case the target field is developmental psychology and psycholinguistics and the relevant research question, how children came to acquire the skill to engage in linguistic negation, e.g. how they acquire the skill to use "no" in order to reject things in appropriate situations and in the appropriate way. For this purpose we implemented minimal cognitive and affective pre-requisites as they were formulated by pragmatic and psycholinguistic theories more than 30 years ago. The humanoid was given a motivation or affect system that was connnected to its symbol grounding system. Its output, a simple trinary value, was treated in terms of symbol grounding in precisely the same manner as if it was yet another sensorial input. The motivation system further triggered the humanoid's facial expression and its bodily behaviour such that the latter was congruent with its motivational state and its facial expression. Equally important, we did not implement any type of logical inference mechanism and neither did we implement any event detection that would operate on temporally extended data.

The construction of this system thus constitutes our first contribution. The system proved to be very successful in the elicitation of negation words in particular, and of intent interpretations, positive or negative, in general. The recorded human-robot interactions show that an approximately natural, linguistically mediated, human-robot interaction is feasible and that this interaction can be effectively exploited in order to contribute to answering long-standing psycholinguistic research questions. The architecture and the constitutive software are available as open source and thus can freely be used by other researchers in order to conduct similar experiments or in order to develop a modified set of behaviours.

Upon construction of the architecture we executed two long-term experiments in order to test two not mutually exclusive hypotheses that address the above mentioned research

question on the origin of negation in language development.

The first hypothesis postulates that it are negative intent interpretations uttered by a caretaker or other adults that constitute the major source of negation words for the child. It is further assumed, equal to the usually tacit assumption of ostensive theories of word acquisition, that these intent interpretations happen simulteously to the child being in the respective emotional or motivational state. A second assumption of this hypothesis is that there is a very high likelihood of negative intent interpretations containing lexical negation words such as "no".

The second hypothesis postulates that it is parental prohibition, i.e. the application of bodily constraints to the child in conjunction with linguistic prohibition that forms the developmental basis of negation. Further assumptions in the context of this hypothesis are (a) that the bodily constraint and the linguistic prohibition are produced simultaneously, (b) that the bodily constraint will lead directly to a negative motivational state on the part of the child, and (c) that it is highly likely that the linguistic prohibitions will contain lexical negation words such as "no", "can't", or "don't".

Based on these two hypotheses and utilizing the developed architecture two blind experiments with naïve participants were conducted. The participants engaged in a linguistically unconstrained conversation with the robot in their stipulated function as language teachers. The participants were not aware of the true purpose of the experiment, but nevertheless produced in both experiments a very high number of negation words compared to the average frequency of negation words in spoken British English as well as compared to a very similar experiment, that utilized the identical humanoid, close to identical experimental instructions, but no motivation system. As a result of the two experiments, none of the two hypotheses could be excluded.

Our second contribution is thus to have produced synthetic, quantitative, and consid-

erably more detailed support for two developmental hypotheses that so far had mainly been descriptive in nature. We have demonstrated in a very concrete and technical way, that both of these potential 'sources of meaning' can act as such, and might constitute alternatives to ostensive theories of meaning.

A third contribution is the advancement of symbol grounding. We have shown that it is principally possible to ground symbols not only in sensorimotor data, but also in emotional or motivational data.

A fourth contribution is the development of negation taxonomies for both negative robot utterances and negative human utterances.

6.3 Discussion of impact

This thesis constitutes the first successful attempt to extend symbol grounding beyond sensorimotor data to include emotional or motivational data. This opens a new avenue of research, in that new kinds of words may be grounded and may be made meaningful for the machine that have so far been out of the reach of robotic language acquisition. The reason that the grounding of these words has not been tackled earlier may have been the, in our opinion doubtful, conviction that they were considered too 'abstract', or only indirectly groundable via other 'more concrete' words. We are thus convinced that negation words are only one group of words that are amenable to such treatment, with emotion words being an obvious second candidate group.

Albeit the idea of grounding symbols or words in anything else than sensorimotor data may seem exotic at first glance, its plausibility is indicated by recent research in experimental psychology that took place in parallel to the work reported in this thesis and unbeknownst to the author (cf. Kousta et al. 2011).

This extension of symbol grounding has allowed us to exploit the apparently natural tendency of humans to engage in intent interpretations when faced with a comparatively incompetent conversation partner, in order to extract and ground words directly from the interaction without the need for third objects or shared attention, but with a heavy reliance on *simultaneity*. Apart from constituting a novel approach to symbol grounding, this interlacing of natural human-robot interaction with a language-centered machine learning approach has the potential to build a bridge between two scientific fields, which so far existed by and large in isolation from each other: affective computing and developmental robotics.

6.4 Future work

As we have indicated earlier, we have gathered more data within the reported experiments than we had opportunity to analyse. First, every participant filled out a *Ten Item Personality Measure (TIPI)* form (Gosling et al. 2003), which would give us an idea of certain personality features of the participants. As the talkativeness but also the style of speech varied considerably between participants, an obvious question is whether or not speech behaviour is somehow correlated with personality features.

Albeit we conducted extensive quantitative and pragmatic analyses on our data, one important analysis is still missing: a full-fledged conversation analysis. Like Fischer et al. (2012) we retrieved some indicators that could automatically be lifted from the data, such as supposed indicators of the speaker's high degree of involvement with the robot. Nonetheless we do consider this method only a shortcut to a proper conversation analysis. We do know as a matter of fact that at least some participants sometimes made their interpretation of the robot's behaviour 'hearable' and therefore observable during the interaction.

Yet so far our examples of such interpretations are of a merely anecdotal kind. We expect to gain far deeper insight into the 'conversational minds' of our participants via a full-fleged conversation analysis (*CA*). Given that the total duration of all 100 sessions amounts to between 8 and 9 hours of video recordings, a full-fledged CA amounts to a considerable effort. Yet we don't see any alternative, if we want to gain a deeper understanding of the precise dynamics of the interaction.

A third avenue of future research is a deeper analysis of the dynamics between the memory-based learner, the robot's utterances, and the participant's linguistic (re-)actions. One could for example envision 'tagging' particular words in the lexicon from the time when they enter the lexicon to the time that 'they' are uttered, similar to the use of radioactive markers in medicine. This most probably requires a modification of the implementation of TiMBL as we are not aware that it would allow such markings. Yet, in order to understand the language acquisition process, one cannot ignore the dynamics between the core learning mechanism and the interactive factors.

Finally, we expect to reach very quickly the limits of what can be done with a purely associative learning mechanism. It would be interesting to take speech act theory even more seriously and couple the grounding mechanism, and/or the motivation module with a goal-evaluation mechanism and possibly switch to some kind of reinforcement learning.

Appendix A

Related Publications

The majority of the work reported within this thesis has not been published yet. Two conference publications explain the theoretical background.

Frank Förster, Chrystopher L. Nehaniv, Joe Saunders, (2011), Robots That Say 'No'. Published in: In George Kampis, István, Eörs Szathmáry, Advances in Artificial Life - Darwin Meets von Neumann, Lecture Notes in Computer Science. Springer. A summary of the negation taxonomies of Bloom, Pea, and Choi and a additional derivation of cognitive requirements for the symbol grounding of negation.

Frank Förster, Chrystopher L. Nehaniv, (2010), Semiotics as Theoretical Underpinning for Language Acquisition in Developmental Robotics. In Klaus Mainzer (ed.), ECAP '10, Proceedings of the VIII European Conference on Computing and Philosophy, TU München, Oct 4-6, 2010. Verlag Dr. Hut. Explains why the speaker cannot be removed from symbol grounding and language acquisition systems and promotes the idea that semiotics, in this context, is a better suited framework than truth-functional semantics.

Appendix B

Additional Tables and Coding Scheme

B.1 Overview of the mapping of lexical negation words to their phonetic counterparts

Table B.1 shows the mapping of lexical negation words to their phonetic counterparts. For those lexical negation words with more than one phonetic counterpart, table B.2 gives an overview of how these different phonetic forms are distributed across participants.

Table B.1: ***List of mappings between lexical negation words and their phonetic counterparts***

lexical	phonetic	lexical	phonetic	lexical	phonetic
no	NOW	doesn't	DAHZAXNT	mustn't	MAASAXNT
don't	DOWNT		DAHZNT		MAHSAXNT
	DUHNT	isn't	IHZAXNT	weren't	WERNT
not	NOHT	haven't	HHAEVAXNT	won't	WOWNT
can't	KAANT	hasn't	HHAEZAXNT	wasn't	WOHZAXNT
	KAENT	couldn't	KUHDAXNT	wouldn't	WUHDAXNT
didn't	DIHDAXNT	neither	NAYDHAX	nono	NOHNOH
	DIHDNT	cannot	KAENAXT		

Table B.2: ***Distribution of phonetic variants for negation words with more than one phonetic variant***

(a) *Rejection Experiment*

	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12
DOWNT	1	26	25	40	45	17	28	2	12	2
DUHNT	1	0	0	0	19	15	17	0	8	6
KAANT	0	0	0	0	0	0	0	0	1	0
KAENT	0	0	0	0	1	0	1	0	1	0
DIHDAXNT	0	5	1	0	0	0	4	0	1	0
DIHDNT	0	0	0	0	0	0	3	0	6	0
DAHZAXNT	0	0	0	1	0	0	1	0	0	0
DAHZNT	0	0	0	0	1	1	0	0	0	0
IHZAXNT	0	0	6	0	2	0	2	0	1	0
IHZNT	0	0	0	0	0	0	0	0	0	0
MAASAXNT	0	0	0	0	0	0	0	0	0	0
MAHSAXNT	0	0	0	0	0	0	0	0	0	0

(b) *Prohibition Experiment*

	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
DOWNT	20	1	49	57	4	27	16	29	21	2
DUHNT	0	0	0	0	0	0	0	0	0	0
KAANT	12	6	19	17	3	13	18	7	0	2
KAENT	0	0	0	0	0	0	0	0	0	0
DIHDAXNT	0	0	4	6	0	1	0	7	0	0
DIHDNT	0	0	0	0	0	0	0	0	0	0
DAHZAXNT	0	0	1	1	1	5	1	0	0	0
DAHZNT	0	0	0	0	0	0	0	0	0	0
IHZAXNT	5	0	0	3	0	7	1	1	0	3
IHZNT	0	0	0	0	0	0	0	0	0	0
MAASAXNT	0	0	0	0	0	0	0	0	0	0
MAHSAXNT	3	0	0	1	1	0	0	0	0	0

B.2 Per session utterance-level measures for speech from Saunders' participants

Table B.3: ***Utterance-level measures for participants speech from Saunders et al. (2012).*** *Any given number refers to the participant with participant id noted on top the corresponding column and the session number in the corresponding first column. Abbreviations: sX: session nr. X, # w/# u: total number of words/utterances uttered by participant, # dw: number of distinct words, MLU: mean length of utterance, w/min / u/min: average number of words / utterances per minute, n/a: data for corresponding session was not available.*

		M02	F05	M03	F01	F02	M01	F03	F06	F04
s1	d (s)	172.7	185.5	170.1	107.1	115.2	167.9	n/a	177.3	120.2
	# w	156	268	120	130	267	210	n/a	371	290
	# u	51	80	41	34	55	62	n/a	99	75
	# dw	29	70	34	42	74	58	n/a	103	85
	MLU	3.1	3.4	2.9	3.8	4.9	3.4	n/a	3.7	3.9
	w/min	54.2	86.7	42.3	72.8	139.1	75.1	n/a	125.5	144.8
	u/min	17.7	25.9	14.5	19	28.6	22.2	n/a	33.5	37.5
s2	d (s)	102.4	130.6	118.9	117.9	125.5	136.9	130.7	138.7	119.7
	# w	105	205	142	145	249	219	214	264	178
	# u	34	48	45	35	55	63	61	77	60
	# dw	24	77	37	37	65	44	56	83	41
	MLU	3.1	4.3	3.2	4.1	4.5	3.5	3.5	3.4	3
	w/min	61.5	94.2	71.7	73.8	119	95.9	98.3	114.2	89.2
	u/min	19.9	22.1	22.7	17.8	26.3	27.6	28	33.3	30.1
s3	d (s)	119.3	129.7	115.5	114	122.7	129	123.3	133.7	128.4
	# w	99	215	97	123	236	162	200	278	220
	# u	31	57	36	34	52	42	64	73	69
	# dw	21	57	39	27	57	36	67	79	61
	MLU	3.2	3.8	2.7	3.6	4.5	3.9	3.1	3.8	3.2
	w/min	49.8	99.4	50.4	64.8	115.4	75.3	97.3	124.7	102.8
	u/min	15.6	26.4	18.7	17.9	25.4	19.5	31.1	32.8	32.2
s4	d (s)	125.9	118.3	116.4	115.8	122.5	126.8	116.7	128.5	126.5
	# w	90	174	82	107	172	127	200	273	192
	# u	35	40	31	28	39	38	61	70	60
	# dw	18	44	29	29	43	25	66	86	52
	MLU	2.6	4.3	2.6	3.8	4.4	3.3	3.3	3.9	3.2
	w/min	42.9	88.2	42.3	55.4	84.3	60.1	102.8	127.5	91
	u/min	16.7	20.3	16	14.5	19.1	18	31.4	32.7	28.4
s5	d (s)	205.7	107.4	125.2	113.8	117.7	102.5	130.7	128.6	126.5
	# w	160	182	99	122	200	93	233	234	155
	# u	53	47	35	28	43	32	76	67	59
	# dw	24	48	37	29	45	23	57	83	46
	MLU	3	3.9	2.8	4.4	4.7	2.9	3.1	3.5	2.6
	w/min	46.7	101.6	47.4	64.3	102	54.5	106.9	109.2	73.5
	u/min	15.5	26.2	16.8	14.8	21.9	18.7	34.9	31.3	28

Table B.4: ***Utterance-level measures for negative utterances of participants speech from Saunders et al. (2012)****. Any given number refers to the participant with participant id noted on top the corresponding column and the session number in the corresponding first column. Abbreviations: sX: session nr. X, # nw/# nu: total number of negative words/utterances uttered by participant, # dnw: number of distinct negative words, MLU: mean length of negative utterances, nw/min / nu/min: average number of neg. words / neg. utterances per minute, n/a: data for corresponding session was not available.*

		M02	F05	M03	F01	F02	M01	F03	F06	F04
s1	d (s)	172.7	185.5	170.1	107.1	115.2	167.9	n/a	177.3	120.2
	# nw	0	0	2	0	0	0	n/a	11	2
	# nu	0	0	1	0	0	0	n/a	10	2
	# dnw	0	0	2	0	0	0	n/a	5	1
	MLU	0	0	19	0	0	0	n/a	6.4	10
	nw/min	0	0	0.7	0	0	0	n/a	3.7	1
	nu/min	0	0	0.4	0	0	0	n/a	3.4	1
s2	d (s)	102.4	130.6	118.9	117.9	125.5	136.9	130.7	138.7	119.7
	# nw	0	0	1	5	0	0	0	14	1
	# nu	0	0	1	4	0	0	0	13	1
	# dnw	0	0	1	2	0	0	0	6	1
	MLU	0	0	3	5.8	0	0	0	5.6	6
	nw/min	0	0	0.5	2.5	0	0	0	6.1	0.5
	nu/min	0	0	0.5	2	0	0	0	5.6	0.5
s3	d (s)	119.3	129.7	115.5	114	122.7	129	123.3	133.7	128.4
	# nw	0	1	3	2	0	0	0	14	3
	# nu	0	1	3	2	0	0	0	13	2
	# dnw	0	1	2	1	0	0	0	5	1
	MLU	0	8	3.7	5	0	0	0	3	2
	nw/min	0	0.5	1.6	1.1	0	0	0	6.3	1.4
	nu/min	0	0.5	1.6	1.1	0	0	0	5.8	0.9
s4	d (s)	125.9	118.3	116.4	115.8	122.5	126.8	116.7	128.5	126.5
	# nw	0	0	2	0	0	0	0	14	0
	# nu	0	0	2	0	0	0	0	12	0
	# dnw	0	0	1	0	0	0	0	6	0
	MLU	0	0	2.5	0	0	0	0	6.6	0
	nw/min	0	0	1	0	0	0	0	6.5	0
	nu/min	0	0	1	0	0	0	0	5.6	0
s5	d (s)	205.7	107.4	125.2	113.8	117.7	102.5	130.7	128.6	126.5
	# nw	0	0	1	1	0	0	1	7	0
	# nu	0	0	1	1	0	0	1	6	0
	# dnw	0	0	1	1	0	0	1	4	0
	MLU	0	0	4	4	0	0	5	6.5	0
	nw/min	0	0	0.5	0.5	0	0	0.5	3.3	0
	nu/min	0	0	0.5	0.5	0	0	0.5	2.8	0

B.3 Complete listings of word frequencies

Table B.5: ***Complete word-frequencies of all words in rejection experiment.*** *Listed are the rank, word count (*cnt*) and the percentage relative to the total number of words in the experiment across all participants and sessions.*

rank	word	cnt	%	rank	word	cnt	%	rank	word	cnt	%
(1)	yuw	1245	7.13	(28)	diyjhiy	127	0.73	(58)	rihmehmbax	40	0.23
(2)	dhax	983	5.63	(29)	dhaets	124	0.71	(58)	hhaxlow	40	0.23
(3)	layk	579	3.31	(30)	uhkey	120	0.69	(59)	sheyp	39	0.22
(4)	ax	475	2.72	(31)	noht	118	0.68	(60)	aem	38	0.22
(5)	dhihs	471	2.7	(32)	hhia	116	0.66	(60)	hhaepih	38	0.22
(6)	now	417	2.39	(33)	wayt	113	0.65	(60)	nays	38	0.22
(7)	wahn	396	2.27	(34)	aol	110	0.63	(61)	feyvaxriht	37	0.21
(8)	skwea	337	1.93	(34)	bohks	110	0.63	(62)	taagiht	36	0.21
(8)	duw	337	1.93	(35)	diychiy	103	0.59	(63)	puht	35	0.2
(9)	tuw	311	1.78	(36)	tray	101	0.58	(63)	uwm	35	0.2
(9)	iht	311	1.78	(37)	yea	99	0.57	(63)	uhey	35	0.2
(10)	dhaet	302	1.73	(37)	trayaenggaxlz	99	0.57	(63)	ihf	35	0.2
(11)	muwn	283	1.62	(38)	hhaev	98	0.56	(64)	shael	32	0.18
(12)	hhaat	279	1.6	(38)	ihn	98	0.56	(64)	fao	32	0.18
(13)	ihz	256	1.47	(39)	vehrih	95	0.54	(65)	dhehn	31	0.18
(14)	trayaenggaxl	254	1.45	(40)	luhk	91	0.52	(66)	hhaats	30	0.17
(15)	serkaxl	231	1.32	(41)	wihdh	90	0.52	(66)	dawn	30	0.17
(16)	ihts	213	1.22	(42)	gow	88	0.5	(66)	yua	30	0.17
(17)	downt	200	1.15	(43)	aet	78	0.45	(66)	ihm	30	0.17
(18)	siy	195	1.12	(44)	yao	77	0.44	(66)	meyk	30	0.17
(18)	aend	194	1.11	(45)	ohv	75	0.43	(67)	wuhd	29	0.17
(18)	wiy	194	1.11	(46)	dhehm	73	0.42	(67)	dhowz	29	0.17
(19)	wehl	193	1.1	(47)	duhnt	69	0.4	(67)	sehntax	29	0.17
(20)	serkaxlz	190	1.09	(47)	goht	69	0.4	(68)	smaol	28	0.16
(21)	ow	187	1.07	(48)	sey	66	0.38	(68)	yey	28	0.16
(22)	skweaz	180	1.03	(49)	lehts	63	0.36	(69)	sahm	27	0.15
(23)	yehs	179	1.02	(50)	sow	59	0.34	(69)	ahrtiyn	27	0.15
(24)	wohnt	172	0.98	(51)	krehsaxnt	58	0.33	(70)	ao	26	0.15
(25)	aa	164	0.94	(51)	ohn	58	0.33	(71)	aez	25	0.14
(26)	dhea	159	0.91	(52)	waots	55	0.31	(72)	feys	24	0.14
(27)	dahn	155	0.89	(53)	gowihng	52	0.3	(73)	er	23	0.13
(28)	woht	153	0.88	(53)	dia	52	0.3	(73)	thriy	23	0.13
(28)	rayt	153	0.88	(52)	thihngk	50	0.29	(73)	mao	23	0.13
(27)	guhd	152	0.87	(52)	hhaw	50	0.29	(72)	shao	22	0.13
(27)	axbawt	152	0.87	(53)	baht	47	0.27	(71)	wahnz	20	0.11
(26)	ay	150	0.86	(54)	axnahdhax	45	0.26	(71)	dhey	20	0.11
(26)	kaen	150	0.86	(55)	naw	44	0.25	(72)	jhahst	19	0.11
(25)	blaek	145	0.83	(56)	saed	43	0.25	(72)	show	19	0.11
(26)	hhowld	135	0.77	(57)	wohts	42	0.24	(73)	miy	18	0.1
(27)	axgehn	133	0.76	(58)	baek	40	0.23	(73)	know	18	0.1

Table B.5 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(74)	taxdey	17	0.1	(81)	biy	10	0.06	(85)	kyuwb	6	0.03
(74)	laykt	17	0.1	(81)	prihtih	10	0.06	(85)	baed	6	0.03
(74)	klehvax	17	0.1	(81)	fayn	10	0.06	(85)	thaot	6	0.03
(74)	dhiyz	17	0.1	(81)	lohts	10	0.06	(86)	naastih	5	0.03
(75)	sheyps	16	0.09	(81)	bihg	10	0.06	(86)	poynts	5	0.03
(75)	hhaez	16	0.09	(82)	laast	9	0.05	(86)	fayv	5	0.03
(75)	wohnax	16	0.09	(82)	geht	9	0.05	(86)	hhaend	5	0.03
(75)	ahp	16	0.09	(82)	sohrih	9	0.05	(86)	yuwaxd	5	0.03
(75)	ihnsayd	16	0.09	(82)	wihl	9	0.05	(86)	nehvax	5	0.03
(75)	taym	16	0.09	(82)	boy	9	0.05	(86)	trayihng	5	0.03
(75)	dihd	16	0.09	(82)	pihraxmihd	9	0.05	(86)	aolsow	5	0.03
(75)	pley	16	0.09	(82)	axwey	9	0.05	(86)	ehnih	5	0.03
(75)	rialih	16	0.09	(83)	ehls	8	0.05	(86)	smayliy	5	0.03
(75)	sayd	16	0.09	(83)	wiyv	8	0.05	(86)	uhps	5	0.03
(75)	ahdhax	16	0.09	(83)	dhow	8	0.05	(86)	stohp	5	0.03
(76)	meybiy	15	0.09	(83)	bawt	8	0.05	(86)	ehs	5	0.03
(76)	mihdaxl	15	0.09	(83)	aydhax	8	0.05	(86)	nehkst	5	0.03
(76)	dahz	15	0.09	(83)	uw	8	0.05	(86)	aolweyz	5	0.03
(77)	thaengk	14	0.08	(83)	teyk	8	0.05	(86)	loht	5	0.03
(77)	gihv	14	0.08	(83)	owvax	8	0.05	(86)	aant	5	0.03
(78)	mehnih	13	0.07	(83)	wer	8	0.05	(86)	hhohraxbaxl	5	0.03
(78)	yuwv	13	0.07	(84)	lahvlih	7	0.04	(86)	rahnihng	5	0.03
(78)	smayl	13	0.07	(84)	rohng	7	0.04	(86)	may	5	0.03
(78)	bohksihz	13	0.07	(84)	hhaa	7	0.04	(86)	blohks	5	0.03
(79)	seyihng	12	0.07	(84)	dheaz	7	0.04	(87)	tayaxd	4	0.02
(79)	wohz	12	0.07	(84)	muwnz	7	0.04	(87)	owld	4	0.02
(79)	yuwax	12	0.07	(84)	lahv	7	0.04	(87)	aen	4	0.02
(79)	hhm	12	0.07	(84)	saydz	7	0.04	(87)	mihlihmiytax	4	0.02
(79)	wihch	12	0.07	(84)	wey	7	0.04	(87)	rehktaenggaxl	4	0.02
(79)	mahch	12	0.07	(84)	awt	7	0.04	(87)	nayt	4	0.02
(79)	biht	12	0.07	(84)	bihkohz	7	0.04	(87)	wiyl	4	0.02
(79)	rehktaenggaxlz	12	0.07	(84)	kwayt	7	0.04	(87)	hhaxm	4	0.02
(79)	meyks	12	0.07	(84)	kiyp	7	0.04	(87)	lahvz	4	0.02
(80)	aym	11	0.06	(84)	werd	7	0.04	(87)	liyv	4	0.02
(80)	ihzaxnt	11	0.06	(84)	skay	7	0.04	(87)	ferst	4	0.02
(80)	wia	11	0.06	(84)	wahrih	7	0.04	(87)	gohn	4	0.02
(80)	hhiaz	11	0.06	(84)	wehn	7	0.04	(87)	piypaxl	4	0.02
(80)	iych	11	0.06	(84)	awax	7	0.04	(87)	rawnd	4	0.02
(80)	gohnax	11	0.06	(84)	owkey	7	0.04	(87)	sahmthihng	4	0.02
(80)	ihl	11	0.06	(84)	kaold	7	0.04	(87)	ehvrihbohdih	4	0.02
(80)	stihl	11	0.06	(84)	siym	7	0.04	(87)	tohp	4	0.02
(80)	way	11	0.06	(84)	sehd	7	0.04	(87)	tiy	4	0.02
(80)	dihdaxnt	11	0.06	(85)	tehl	6	0.03	(87)	ahdhaxz	4	0.02
(81)	dihdnt	10	0.06	(85)	sihks	6	0.03	(87)	yuwzhaxlih	4	0.02
(81)	ayl	10	0.06	(85)	aekchualih	6	0.03	(87)	smaolax	4	0.02
(81)	hhaaf	10	0.06	(85)	luhks	6	0.03	(87)	poyntiy	4	0.02

Table B.5 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(87)	feyvaxrihts	4	0.02	(88)	seym	3	0.02	(89)	mahdaxld	2	0.01
(87)	mayt	4	0.02	(88)	mawdh	3	0.02	(89)	thihngz	2	0.01
(87)	iyvaxn	4	0.02	(88)	hhaatih	3	0.02	(89)	prohbaxblih	2	0.01
(87)	dihfraxnt	4	0.02	(88)	bay	3	0.02	(89)	ihndihfraxnt	2	0.01
(87)	luhkihng	4	0.02	(88)	liyst	3	0.02	(89)	sheypt	2	0.01
(87)	behtax	4	0.02	(88)	baekgrawnd	3	0.02	(89)	saend	2	0.01
(87)	hhaed	4	0.02	(88)	kiyn	3	0.02	(89)	sahmtaymz	2	0.01
(87)	axrawnd	4	0.02	(88)	grahmpih	3	0.02	(89)	blohk	2	0.01
(87)	maysehlf	4	0.02	(88)	ayv	3	0.02	(89)	fyuw	2	0.01
(87)	pleyihng	4	0.02	(88)	miyn	3	0.02	(89)	jhohb	2	0.01
(87)	hhowldihng	4	0.02	(88)	kaoz	3	0.02	(89)	wohntihd	2	0.01
(87)	hheyt	4	0.02	(88)	hhey	3	0.02	(89)	saxrawndihd	2	0.01
(87)	ihnahf	4	0.02	(88)	maynd	3	0.02	(89)	pliyz	2	0.01
(87)	ahpsayd	4	0.02	(88)	kaol	3	0.02	(89)	lernihng	2	0.01
(87)	weyk	4	0.02	(88)	ahnyuwzhaxl	3	0.02	(89)	leht	2	0.01
(87)	dyuh	4	0.02	(88)	sheykihng	3	0.02	(89)	tern	2	0.01
(87)	smaylihng	4	0.02	(88)	yuwd	3	0.02	(89)	faen	2	0.01
(87)	faynd	4	0.02	(88)	hhehd	3	0.02	(89)	kaxnsehntrihk	2	0.01
(87)	yeht	4	0.02	(89)	gehtihng	2	0.01	(89)	fahn	2	0.01
(87)	laajh	4	0.02	(89)	yuwst	2	0.01	(89)	faekt	2	0.01
(87)	uh	4	0.02	(89)	kaos	2	0.01	(89)	streytaxwey	2	0.01
(87)	mowst	4	0.02	(89)	ehvrih	2	0.01	(89)	th	2	0.01
(87)	hhaevaxnt	4	0.02	(89)	ehnihwahn	2	0.01	(89)	kaynd	2	0.01
(88)	kaent	3	0.02	(89)	rihd	2	0.01	(89)	wuhm	2	0.01
(88)	throwihng	3	0.02	(89)	axpihnianz	2	0.01	(89)	yuwl	2	0.01
(88)	behst	3	0.02	(89)	prohmihs	2	0.01	(89)	axweyk	2	0.01
(88)	ayf	3	0.02	(89)	layn	2	0.01	(89)	throw	2	0.01
(88)	kaxrehkt	3	0.02	(89)	lowdz	2	0.01	(89)	dhaen	2	0.01
(88)	byuwtaxfuhl	3	0.02	(89)	ehnihwey	2	0.01	(89)	gowz	2	0.01
(88)	lihtaxl	3	0.02	(89)	paetaxnz	2	0.01	(89)	paat	2	0.01
(88)	dehfihnaxtlih	3	0.02	(89)	kaech	2	0.01	(89)	siyn	2	0.01
(88)	waw	3	0.02	(89)	fahnih	2	0.01	(89)	liy	2	0.01
(88)	ayz	3	0.02	(89)	biyn	2	0.01	(89)	aaftax	2	0.01
(88)	ahndaxstaend	3	0.02	(89)	staat	2	0.01	(89)	sawnd	2	0.01
(88)	niyd	3	0.02	(89)	aebsaxluwtlih	2	0.01	(89)	kahlax	2	0.01
(88)	axgehnst	3	0.02	(89)	ownlih	2	0.01	(89)	axpaeraxntlih	2	0.01
(88)	ruwmz	3	0.02	(89)	ihntraxstihd	2	0.01	(89)	nayslih	2	0.01
(88)	wownt	3	0.02	(89)	toyz	2	0.01	(89)	klows	2	0.01
(88)	ehksaxlaxnt	3	0.02	(89)	maetax	2	0.01	(89)	yaoz	2	0.01
(88)	tihlt	3	0.02	(89)	kohs	2	0.01	(89)	tahch	2	0.01
(88)	wihdhihn	3	0.02	(89)	prihferd	2	0.01	(89)	frohm	2	0.01
(88)	kahmz	3	0.02	(89)	hhayd	2	0.01	(89)	baorihng	2	0.01
(88)	pihraxmihdz	3	0.02	(89)	behdtaym	2	0.01	(89)	drohp	2	0.01
(88)	naa	3	0.02	(89)	siymz	2	0.01	(89)	dahzaxnt	2	0.01
(88)	lern	3	0.02	(89)	kuhd	2	0.01	(89)	dahznt	2	0.01
(88)	pihk	3	0.02	(89)	puhtihng	2	0.01	(89)	drohpt	2	0.01

Table B.5 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(89)	ehm	2	0.01	(90)	meyd	1	0.01	(90)	yuwzhaxl	1	0.01
(89)	wiyk	2	0.01	(90)	baelaxnsihng	1	0.01	(90)	dey	1	0.01
(89)	ohbviaslih	2	0.01	(90)	sehvraxl	1	0.01	(90)	taymz	1	0.01
(89)	aaskihng	2	0.01	(90)	dihslayk	1	0.01	(90)	kuwl	1	0.01
(89)	wiyd	2	0.01	(90)	wihndowz	1	0.01	(90)	aam	1	0.01
(89)	sihlih	2	0.01	(90)	saot	1	0.01	(90)	sawndihng	1	0.01
(89)	staek	2	0.01	(90)	kaolihng	1	0.01	(90)	bihld	1	0.01
(89)	siarias	2	0.01	(90)	axm	1	0.01	(90)	md	1	0.01
(90)	sihlvax	1	0.01	(90)	kwihk	1	0.01	(90)	muwd	1	0.01
(90)	trayd	1	0.01	(90)	wea	1	0.01	(90)	saakaezaxm	1	0.01
(90)	ahpseht	1	0.01	(90)	kaot	1	0.01	(90)	taagihts	1	0.01
(90)	poyntiyah	1	0.01	(90)	werkt	1	0.01	(90)	aansax	1	0.01
(90)	shuhd	1	0.01	(90)	ohpshaxnz	1	0.01	(90)	paast	1	0.01
(90)	kihdz	1	0.01	(90)	dhaw	1	0.01	(90)	sahch	1	0.01
(90)	praektihs	1	0.01	(90)	kahlaxz	1	0.01	(90)	siyihng	1	0.01
(90)	aembihvaxlaxnt	1	0.01	(90)	frawniy	1	0.01	(90)	miynz	1	0.01
(90)	ihntraxst	1	0.01	(90)	ihksehl	1	0.01	(90)	ihksprehs	1	0.01
(90)	ternihng	1	0.01	(90)	hhaendz	1	0.01	(90)	growihng	1	0.01
(90)	ihntraxstihng	1	0.01	(90)	faynaxlih	1	0.01	(90)	braek	1	0.01
(90)	fayax	1	0.01	(90)	hhay	1	0.01	(90)	werk	1	0.01
(90)	aagyuhmehntaxtihv	1	0.01	(90)	ey	1	0.01	(90)	aolrehdih	1	0.01
(90)	taok	1	0.01	(90)	aagyuw	1	0.01	(90)	maaksmaxn	1	0.01
(90)	taa	1	0.01	(90)	stey	1	0.01	(90)	neymz	1	0.01
(90)	tehraxblih	1	0.01	(90)	waa	1	0.01	(90)	kayndax	1	0.01
(90)	tayp	1	0.01	(90)	wuhdaxnt	1	0.01	(90)	chuwz	1	0.01
(90)	ihfehkts	1	0.01	(90)	taynih	1	0.01	(90)	kawnt	1	0.01
(90)	z	1	0.01	(90)	gehst	1	0.01	(90)	lahvd	1	0.01
(90)	mkey	1	0.01	(90)	aydia	1	0.01	(90)	thaengks	1	0.01
(90)	sao	1	0.01	(90)	kaxnehktihd	1	0.01	(90)	ehnihthihng	1	0.01
(90)	pehn	1	0.01	(90)	hhiy	1	0.01	(90)	ihmaejhihn	1	0.01
(90)	hhaevihng	1	0.01	(90)	hhowp	1	0.01	(90)	baekwaxdz	1	0.01
(90)	poynt	1	0.01	(90)	dihprehst	1	0.01	(90)	hhaws	1	0.01
(90)	luhkt	1	0.01	(90)	axlawd	1	0.01	(90)	cheynjh	1	0.01
(90)	kervz	1	0.01	(90)	kuhdaxnt	1	0.01	(90)	ohf	1	0.01
(90)	vaelaxntaynz	1	0.01	(90)	bluw	1	0.01	(90)	duwihng	1	0.01
(90)	mowmaxnt	1	0.01	(90)	bow	1	0.01	(90)	brihng	1	0.01
(90)	bohtaxm	1	0.01	(90)	meykihng	1	0.01	(90)	yeyy	1	0.01
(90)	ehl	1	0.01	(90)	wihndow	1	0.01	(90)	slow	1	0.01
(90)	puhshihng	1	0.01	(90)	saxprayzd	1	0.01	(90)	ihnsaydz	1	0.01
(90)	breyk	1	0.01	(90)	suwn	1	0.01	(90)	tehraxbaxl	1	0.01
(90)	striyt	1	0.01	(90)	raadhax	1	0.01	(90)	sohft	1	0.01
(90)	sliyp	1	0.01	(90)	guhdnaxs	1	0.01	(90)	prihfer	1	0.01
(90)	wowkaxn	1	0.01	(90)	vehrihg	1	0.01	(90)	lernihd	1	0.01
(90)	sihmbaxlz	1	0.01	(90)	spowk	1	0.01	(90)	faos	1	0.01
(90)	bihfao	1	0.01	(90)	shehl	1	0.01	(90)	hhah	1	0.01
(90)	fohnd	1	0.01	(90)	fyuwchax	1	0.01	(90)	fea	1	0.01

Table B.5 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(90)	ehniymao	1	0.01	(90)	raytuh	1	0.01	(90)	doht	1	0.01
(90)	neym	1	0.01	(90)	ahhhah	1	0.01	(90)	low	1	0.01
(90)	kriypih	1	0.01	(90)	niychiy	1	0.01	(90)	paetaxn	1	0.01
(90)	rehktaenggyuhlax	1	0.01	(90)	axnoyd	1	0.01	(90)	kahm	1	0.01
(90)	layf	1	0.01	(90)	iychiy	1	0.01	(90)	deyz	1	0.01
(90)	wohzaxnt	1	0.01	(90)	ihnstehd	1	0.01	(90)	dihsaysihv	1	0.01
(90)	flao	1	0.01	(90)	gehs	1	0.01	(90)	daansihng	1	0.01
(90)	ihksehpt	1	0.01	(90)	sihmbaxl	1	0.01	(90)	naomaxlih	1	0.01
(90)	rihmaxm	1	0.01	(90)	spoht	1	0.01	(90)	ohlrayt	1	0.01
(90)	ehkspert	1	0.01	(90)	bihhhaynd	1	0.01	(90)	yuws	1	0.01
(90)	ihgnao	1	0.01	(90)	ihnkohmpihtaxnt	1	0.01	(90)	paxtihkyuhlaxlih	1	0.01
(90)	uhd	1	0.01	(90)	rihaektihng	1	0.01	(90)	mhhm	1	0.01
(90)	shown	1	0.01	(90)	ahnhhaepih	1	0.01	(90)	mihstriytihng	1	0.01
(90)	ael	1	0.01	(90)	piypaxlz	1	0.01	(90)	kaxrehktihd	1	0.01
(90)	wearaez	1	0.01	(90)	wayts	1	0.01	(90)	sayn	1	0.01
(90)	eywihs	1	0.01	(90)	wahns	1	0.01	(90)	mehnt	1	0.01
(90)	dheyv	1	0.01	(90)	ihnkraed	1	0.01	(90)	nyuw	1	0.01
(90)	fihnihsht	1	0.01	(90)	fiylihng	1	0.01	(90)	kaant	1	0.01
(90)	laykihng	1	0.01	(90)	ihmpaotaxnt	1	0.01	(90)	yuwaxl	1	0.01
(90)	rowlihng	1	0.01	(90)	ehsey	1	0.01	(90)	hheytihd	1	0.01
(90)	hhow	1	0.01	(90)	wihsh	1	0.01	(90)	hhehld	1	0.01
(90)	sax	1	0.01	(90)	laots	1	0.01	(90)	gohtax	1	0.01
(90)	ihgzaektlih	1	0.01	(90)	frawn	1	0.01	(90)	wernt	1	0.01
(90)	mihzaxraxbaxl	1	0.01	(90)	paenihkihng	1	0.01	(90)	ihnkrehdaxbaxl	1	0.01
(90)	gerl	1	0.01	(90)	aolmowst	1	0.01	(90)	stahk	1	0.01
(90)	ihntax	1	0.01								

Table B.6: ***Complete word-frequencies of salient words in rejection experiment.*** *Listed are the rank, word count (*cnt*) and the percentage relative to the total number of words in the experiment across all participants and sessions.*

rank	word	cnt	%	rank	word	cnt	%	rank	word	cnt	%
(1)	skwea	259	4.97	(35)	dhehm	25	0.48	(47)	thihngk	11	0.21
(2)	now	242	4.64	(35)	ahrtiyn	25	0.48	(47)	hhaxlow	11	0.21
(3)	trayaenggaxl	206	3.95	(35)	wohnt	25	0.48	(48)	sayd	10	0.19
(4)	hhaat	198	3.8	(35)	vehrih	25	0.48	(48)	hhm	10	0.19
(5)	muwn	184	3.53	(36)	tray	24	0.46	(49)	wahnz	9	0.17
(5)	serkaxl	184	3.53	(36)	noht	24	0.46	(49)	dhowz	9	0.17
(6)	layk	167	3.2	(37)	hhaats	22	0.42	(49)	mihdaxl	9	0.17
(7)	serkaxlz	140	2.69	(37)	sehntax	22	0.42	(50)	sey	8	0.15
(8)	skweaz	126	2.42	(38)	dia	20	0.38	(50)	miy	8	0.15
(9)	iht	123	2.36	(38)	lehts	20	0.38	(50)	ohn	8	0.15
(10)	yehs	119	2.28	(38)	tuw	20	0.38	(50)	taym	8	0.15
(11)	wahn	111	2.13	(38)	hhaepih	20	0.38	(50)	pihraxmihd	8	0.15
(12)	rayt	98	1.88	(39)	hhia	19	0.36	(51)	ay	7	0.13
(13)	dhihs	95	1.82	(39)	waots	19	0.36	(51)	baek	7	0.13
(14)	uhkey	90	1.73	(39)	dawn	19	0.36	(51)	hhaw	7	0.13
(15)	axgehn	88	1.69	(39)	luhk	19	0.36	(51)	fao	7	0.13
(16)	ow	82	1.57	(39)	sheyp	19	0.36	(51)	klehvax	7	0.13
(17)	guhd	81	1.55	(39)	nays	19	0.36	(51)	pley	7	0.13
(18)	diyjhiy	76	1.46	(40)	uhey	18	0.35	(51)	fayn	7	0.13
(19)	diychiy	68	1.3	(41)	ihts	17	0.33	(52)	yao	6	0.12
(20)	dahn	63	1.21	(41)	dhax	17	0.33	(52)	thaengk	6	0.12
(21)	dhaet	62	1.19	(42)	axnahdhax	16	0.31	(52)	aydhax	6	0.12
(21)	trayaenggaxlz	62	1.19	(42)	dhaets	16	0.31	(52)	thriy	6	0.12
(22)	gow	58	1.11	(43)	shao	15	0.29	(52)	baht	6	0.12
(23)	yuw	53	1.02	(43)	sow	15	0.29	(52)	prihtih	6	0.12
(23)	axbawt	53	1.02	(44)	ihz	14	0.27	(52)	wiy	6	0.12
(24)	downt	52	1	(44)	naw	14	0.27	(52)	rialih	6	0.12
(25)	siy	47	0.9	(45)	sheyps	13	0.25	(52)	wohts	6	0.12
(25)	hhowld	47	0.9	(45)	woht	13	0.25	(53)	skay	5	0.1
(26)	wayt	46	0.88	(45)	hhaev	13	0.25	(53)	wahrih	5	0.1
(27)	yea	45	0.86	(45)	rihmehmbax	13	0.25	(53)	ihzaxnt	5	0.1
(27)	bohks	45	0.86	(46)	smayl	12	0.23	(53)	ahp	5	0.1
(28)	wehl	41	0.79	(46)	aol	12	0.23	(53)	mahch	5	0.1
(29)	krehsaxnt	39	0.75	(46)	yey	12	0.23	(53)	ahdhax	5	0.1
(30)	dhea	37	0.71	(46)	duhnt	12	0.23	(53)	laast	5	0.1
(31)	aa	36	0.69	(46)	smaol	12	0.23	(53)	owkey	5	0.1
(31)	blaek	36	0.69	(46)	duw	12	0.23	(53)	meyk	5	0.1
(32)	aend	35	0.67	(47)	bohksihz	11	0.21	(53)	smayliy	5	0.1
(32)	saed	35	0.67	(47)	goht	11	0.21	(53)	dhehn	5	0.1
(33)	taagiht	31	0.59	(47)	er	11	0.21	(54)	feyvaxrihts	4	0.08
(33)	uwm	31	0.59	(47)	rehktaenggaxlz	11	0.21	(54)	rehktaenggaxl	4	0.08
(34)	feyvaxriht	27	0.52	(47)	ax	11	0.21	(54)	hhaxm	4	0.08
(34)	aem	27	0.52	(47)	taxdey	11	0.21	(54)	sohrih	4	0.08

Table B.6 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(54)	wihch	4	0.08	(55)	blohks	3	0.06	(56)	ehnih	2	0.04
(54)	stihl	4	0.08	(56)	poyntiy	2	0.04	(56)	meybiy	2	0.04
(54)	biht	4	0.08	(56)	lernihng	2	0.04	(56)	drohpt	2	0.04
(54)	uw	4	0.08	(56)	kaxnsehntrihk	2	0.04	(56)	uh	2	0.04
(54)	dhow	4	0.08	(56)	geht	2	0.04	(56)	gihv	2	0.04
(54)	baed	4	0.08	(56)	ehnihwahn	2	0.04	(56)	maynd	2	0.04
(54)	saydz	4	0.08	(56)	tihlt	2	0.04	(56)	ahnyuwzhaxl	2	0.04
(54)	teyk	4	0.08	(56)	streytaxwey	2	0.04	(56)	saend	2	0.04
(54)	show	4	0.08	(56)	kaent	2	0.04	(56)	wiyk	2	0.04
(54)	aekchualih	4	0.08	(56)	th	2	0.04	(56)	kwayt	2	0.04
(54)	feys	4	0.08	(56)	naastih	2	0.04	(56)	yuwzhaxlih	2	0.04
(54)	bihkohz	4	0.08	(56)	ehls	2	0.04	(56)	ohbviaslih	2	0.04
(54)	mao	4	0.08	(56)	rohng	2	0.04	(56)	ruwmz	2	0.04
(54)	wihdh	4	0.08	(56)	kyuwb	2	0.04	(56)	jhohb	2	0.04
(54)	ahdhaxz	4	0.08	(56)	wuhm	2	0.04	(56)	siarias	2	0.04
(54)	laykt	4	0.08	(56)	ehnihwey	2	0.04	(56)	wownt	2	0.04
(54)	uhps	4	0.08	(56)	axweyk	2	0.04	(57)	pliyz	1	0.02
(54)	lohts	4	0.08	(56)	aym	2	0.04	(57)	sihlvax	1	0.02
(55)	tayaxd	3	0.06	(56)	paetaxnz	2	0.04	(57)	neymz	1	0.02
(55)	ehksaxlaxnt	3	0.06	(56)	wohnax	2	0.04	(57)	kaos	1	0.02
(55)	gowihng	3	0.06	(56)	puht	2	0.04	(57)	trayd	1	0.02
(55)	lahvlih	3	0.06	(56)	poynts	2	0.04	(57)	faen	1	0.02
(55)	throwihng	3	0.06	(56)	ferst	2	0.04	(57)	ahpseht	1	0.02
(55)	dihfraxnt	3	0.06	(56)	staat	2	0.04	(57)	wuhd	1	0.02
(55)	luhkihng	3	0.06	(56)	aebsaxluwtlih	2	0.04	(57)	dihdnt	1	0.02
(55)	pihraxmihdz	3	0.06	(56)	aet	2	0.04	(57)	praektihs	1	0.02
(55)	seyihng	3	0.06	(56)	ayz	2	0.04	(57)	kayndax	1	0.02
(55)	byuwtaxfuhl	3	0.06	(56)	maetax	2	0.04	(57)	aembihvaxlaxnt	1	0.02
(55)	shael	3	0.06	(56)	boy	2	0.04	(57)	kawnt	1	0.02
(55)	sihks	3	0.06	(56)	kaold	2	0.04	(57)	ihntraxstihng	1	0.02
(55)	hhaaf	3	0.06	(56)	bay	2	0.04	(57)	lahvd	1	0.02
(55)	jhahst	3	0.06	(56)	loht	2	0.04	(57)	ihmaejhihn	1	0.02
(55)	dhey	3	0.06	(56)	klows	2	0.04	(57)	faekt	1	0.02
(55)	ihnsayd	3	0.06	(56)	piypaxl	2	0.04	(57)	baekwaxdz	1	0.02
(55)	mawdh	3	0.06	(56)	baekgrawnd	2	0.04	(57)	axpihnianz	1	0.02
(55)	dhiyz	3	0.06	(56)	ahndaxstaend	2	0.04	(57)	wihdhihn	1	0.02
(55)	waw	3	0.06	(56)	grahmpih	2	0.04	(57)	cheynjh	1	0.02
(55)	muwnz	3	0.06	(56)	yeht	2	0.04	(57)	ohf	1	0.02
(55)	hhaatih	3	0.06	(56)	rawnd	2	0.04	(57)	mayt	1	0.02
(55)	ahpsayd	3	0.06	(56)	trayihng	2	0.04	(57)	tehraxblih	1	0.02
(55)	smaylihng	3	0.06	(56)	baorihng	2	0.04	(57)	luhks	1	0.02
(55)	aant	3	0.06	(56)	ihn	2	0.04	(57)	ihfehkts	1	0.02
(55)	hhohraxbaxl	3	0.06	(56)	prohbaxblih	2	0.04	(57)	ayf	1	0.02
(55)	smaolax	3	0.06	(56)	hhey	2	0.04	(57)	yeyy	1	0.02
(55)	sahm	3	0.06	(56)	ehvrihbohdih	2	0.04	(57)	ihl	1	0.02
(55)	axwey	3	0.06	(56)	sehd	2	0.04	(57)	kaxrehkt	1	0.02

Table B.6 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(57)	mkey	1	0.02	(57)	nehkst	1	0.02	(57)	hhayd	1	0.02
(57)	ihnsaydz	1	0.02	(57)	ohpshaxnz	1	0.02	(57)	faynd	1	0.02
(57)	tehraxbaxl	1	0.02	(57)	maysehlf	1	0.02	(57)	kuhdaxnt	1	0.02
(57)	behtax	1	0.02	(57)	shown	1	0.02	(57)	ihmpaotaxnt	1	0.02
(57)	layn	1	0.02	(57)	ihf	1	0.02	(57)	bluw	1	0.02
(57)	nayt	1	0.02	(57)	awax	1	0.02	(57)	behdtaym	1	0.02
(57)	pehn	1	0.02	(57)	pleyihng	1	0.02	(57)	bow	1	0.02
(57)	luhkt	1	0.02	(57)	ihntraxstihd	1	0.02	(57)	tahch	1	0.02
(57)	kervz	1	0.02	(57)	fayv	1	0.02	(57)	wihsh	1	0.02
(57)	hhaa	1	0.02	(57)	kahlaxz	1	0.02	(57)	wey	1	0.02
(57)	vaelaxntaynz	1	0.02	(57)	hheyt	1	0.02	(57)	rahnihng	1	0.02
(57)	mowmaxnt	1	0.02	(57)	fihnihsht	1	0.02	(57)	laots	1	0.02
(57)	bohtaxm	1	0.02	(57)	rowlihng	1	0.02	(57)	frawn	1	0.02
(57)	ohv	1	0.02	(57)	toyz	1	0.02	(57)	paenihkihng	1	0.02
(57)	ehl	1	0.02	(57)	ihnahf	1	0.02	(57)	aolmowst	1	0.02
(57)	faos	1	0.02	(57)	ao	1	0.02	(57)	kuhd	1	0.02
(57)	stohp	1	0.02	(57)	hhow	1	0.02	(57)	doht	1	0.02
(57)	naa	1	0.02	(57)	faynaxlih	1	0.02	(57)	hhiaz	1	0.02
(57)	hhaez	1	0.02	(57)	mihzaxraxbaxl	1	0.02	(57)	saxprayzd	1	0.02
(57)	hhah	1	0.02	(57)	ihgzaektlih	1	0.02	(57)	paetaxn	1	0.02
(57)	hhaed	1	0.02	(57)	hhaend	1	0.02	(57)	puhtihng	1	0.02
(57)	ayl	1	0.02	(57)	hhay	1	0.02	(57)	suwn	1	0.02
(57)	wowkaxn	1	0.02	(57)	ey	1	0.02	(57)	raadhax	1	0.02
(57)	sihmbaxlz	1	0.02	(57)	aolweyz	1	0.02	(57)	ayv	1	0.02
(57)	neym	1	0.02	(57)	weyk	1	0.02	(57)	aolsow	1	0.02
(57)	wehn	1	0.02	(57)	raytuh	1	0.02	(57)	deyz	1	0.02
(57)	kriypih	1	0.02	(57)	ahhhah	1	0.02	(57)	guhdnaxs	1	0.02
(57)	kaech	1	0.02	(57)	niychiy	1	0.02	(57)	sahmthihng	1	0.02
(57)	bihg	1	0.02	(57)	kahlax	1	0.02	(57)	ihndihfraxnt	1	0.02
(57)	rehktaenggyuhlax	1	0.02	(57)	axnoyd	1	0.02	(57)	laajh	1	0.02
(57)	baelaxnsihng	1	0.02	(57)	lahv	1	0.02	(57)	owvax	1	0.02
(57)	axrawnd	1	0.02	(57)	kohs	1	0.02	(57)	tohp	1	0.02
(57)	ehs	1	0.02	(57)	gehst	1	0.02	(57)	daansihng	1	0.02
(57)	sehvraxl	1	0.02	(57)	aydia	1	0.02	(57)	dahzaxnt	1	0.02
(57)	fahnih	1	0.02	(57)	nayslih	1	0.02	(57)	taymz	1	0.02
(57)	dihslayk	1	0.02	(57)	ihnstehd	1	0.02	(57)	dihdaxnt	1	0.02
(57)	layf	1	0.02	(57)	sihmbaxl	1	0.02	(57)	kuwl	1	0.02
(57)	wihndowz	1	0.02	(57)	gehs	1	0.02	(57)	ohlrayt	1	0.02
(57)	ihksehpt	1	0.02	(57)	kaxnehktihd	1	0.02	(57)	bihld	1	0.02
(57)	flao	1	0.02	(57)	bihhhaynd	1	0.02	(57)	meyks	1	0.02
(57)	siyn	1	0.02	(57)	yaoz	1	0.02	(57)	axgehnst	1	0.02
(57)	dahz	1	0.02	(57)	prihferd	1	0.02	(57)	saakaezaxm	1	0.02
(57)	ehkspert	1	0.02	(57)	ihnkohmpihtaxnt	1	0.02	(57)	taagihts	1	0.02
(57)	kaot	1	0.02	(57)	rihaektihng	1	0.02	(57)	aansax	1	0.02
(57)	werkt	1	0.02	(57)	ahnhhaepih	1	0.02	(57)	paast	1	0.02
(57)	dehfihnaxtlih	1	0.02	(57)	wayts	1	0.02	(57)	mhhm	1	0.02

Table B.6 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(57)	sheykihng	1	0.02	(57)	hhehd	1	0.02	(57)	sihlih	1	0.02
(57)	sahmtaymz	1	0.02	(57)	hhaevaxnt	1	0.02	(57)	staek	1	0.02
(57)	sahch	1	0.02	(57)	growihng	1	0.02	(57)	ihnkrehdaxbaxl	1	0.02
(57)	mihstriytihng	1	0.02	(57)	aaskihng	1	0.02	(57)	wer	1	0.02
(57)	ihksprehs	1	0.02	(57)	hhehld	1	0.02	(57)	werk	1	0.02
(57)	mowst	1	0.02	(57)	saxrawndihd	1	0.02	(57)	kiyp	1	0.02

Table B.7: ***Complete word-frequencies of all words in prohibition experiment****. Listed are the rank, word count (*cnt*) and the percentage relative to the total number of words in the experiment across all participants and sessions.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	yuw	1591	6.18	(30)	aa	199	0.77	(60)	goht	91	0.35
(2)	dhax	1416	5.5	(31)	noht	198	0.77	(61)	naw	89	0.35
(3)	ax	962	3.74	(32)	krehsaxnt	184	0.71	(62)	rialih	87	0.34
(4)	dhihs	956	3.71	(32)	axgehn	184	0.71	(63)	skweaz	86	0.33
(5)	wahn	722	2.8	(33)	serkaxlz	168	0.65	(64)	wohnax	82	0.32
(6)	ihz	632	2.46	(34)	ay	167	0.65	(65)	ow	80	0.31
(7)	layk	527	2.05	(35)	woht	165	0.64	(66)	hhaats	79	0.31
(8)	tuw	471	1.83	(36)	wehl	162	0.63	(67)	axnahdhax	78	0.3
(9)	now	461	1.79	(37)	wohnt	159	0.62	(68)	hhaw	74	0.29
(10)	ihts	428	1.66	(38)	axbawt	157	0.61	(69)	trayaenggaxlz	72	0.28
(11)	hhaat	411	1.6	(39)	siy	148	0.57	(70)	kaold	71	0.28
(12)	aend	389	1.51	(40)	lehts	143	0.56	(71)	yea	68	0.26
(12)	skwea	389	1.51	(41)	rawnd	140	0.54	(72)	ihf	66	0.26
(13)	trayaenggaxl	377	1.46	(42)	gow	133	0.52	(72)	tahch	66	0.26
(14)	dhaet	366	1.42	(43)	diychiy	132	0.51	(73)	puht	65	0.25
(14)	iht	366	1.42	(44)	baht	130	0.51	(74)	yao	62	0.24
(15)	duw	360	1.4	(45)	dhea	125	0.49	(75)	aol	59	0.23
(16)	muwn	356	1.38	(45)	rihmehmbax	125	0.49	(76)	nays	59	0.23
(17)	serkaxl	332	1.29	(46)	rayt	124	0.48	(77)	dawn	57	0.22
(18)	dhaets	329	1.28	(47)	bohks	123	0.48	(77)	shael	57	0.22
(19)	sheyp	310	1.2	(48)	yua	121	0.47	(77)	axlawd	57	0.22
(20)	luhk	286	1.11	(49)	sow	116	0.45	(78)	dhey	56	0.22
(20)	aet	286	1.11	(50)	uhey	115	0.45	(79)	sheyps	55	0.21
(21)	wiy	281	1.09	(51)	hhia	113	0.44	(80)	thihngk	54	0.21
(22)	wihdh	269	1.04	(51)	ohn	113	0.44	(81)	ihn	53	0.21
(23)	kaen	266	1.03	(52)	hhowld	111	0.43	(82)	wia	51	0.2
(24)	vehrih	263	1.02	(53)	fao	110	0.43	(82)	saydz	51	0.2
(25)	guhd	258	1	(54)	dahn	106	0.41	(82)	dhehn	51	0.2
(26)	pley	229	0.89	(55)	ohv	104	0.4	(82)	know	51	0.2
(26)	downt	229	0.89	(56)	diyjhiy	103	0.4	(83)	sey	49	0.19
(27)	yehs	227	0.88	(57)	baek	99	0.38	(83)	gowihng	49	0.19
(28)	uhkey	220	0.85	(58)	kaant	98	0.38	(83)	miy	49	0.19
(29)	hhaev	207	0.8	(59)	taxdey	94	0.37	(84)	tray	48	0.19

Table B.7 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(84)	thriy	48	0.19	(103)	thaengk	22	0.09	(112)	er	13	0.05
(85)	aym	44	0.17	(103)	ahdhax	22	0.09	(112)	pleyihng	13	0.05
(85)	wihch	44	0.17	(103)	wihl	22	0.09	(112)	blohks	13	0.05
(86)	taym	42	0.16	(103)	baorihng	22	0.09	(113)	prehzaxnt	12	0.05
(87)	wuhd	40	0.16	(104)	miynz	21	0.08	(113)	hhaaf	12	0.05
(87)	hhaxlow	40	0.16	(104)	smayl	21	0.08	(113)	hhaend	12	0.05
(88)	aez	39	0.15	(105)	ihzaxnt	20	0.08	(113)	waonihng	12	0.05
(88)	ehksaxlaxnt	39	0.15	(105)	aolsow	20	0.08	(113)	tohp	12	0.05
(88)	may	39	0.15	(105)	luhkihng	20	0.08	(113)	gowz	12	0.05
(89)	sohrih	37	0.14	(105)	show	20	0.08	(113)	klehvax	12	0.05
(90)	uwm	35	0.14	(106)	lahvlih	19	0.07	(113)	shaap	12	0.05
(90)	kwayt	35	0.14	(106)	lohts	19	0.07	(113)	loht	12	0.05
(90)	nehkst	35	0.14	(106)	luhks	19	0.07	(113)	siym	12	0.05
(90)	dhiyz	35	0.14	(106)	kyuwb	19	0.07	(113)	aant	12	0.05
(91)	dhehm	34	0.13	(106)	dihdaxnt	19	0.07	(113)	sehd	12	0.05
(91)	feyvaxriht	34	0.13	(107)	kuhd	18	0.07	(114)	bihfao	11	0.04
(91)	gohnax	34	0.13	(107)	awax	18	0.07	(114)	ihntraxstihd	11	0.04
(92)	blaek	33	0.13	(107)	biht	18	0.07	(114)	aekchualih	11	0.04
(93)	teyk	32	0.12	(108)	biy	17	0.07	(114)	awt	11	0.04
(94)	lahv	31	0.12	(108)	axrawnd	17	0.07	(114)	skay	11	0.04
(94)	jhahst	31	0.12	(108)	mahch	17	0.07	(114)	bihldihng	11	0.04
(95)	wiyl	30	0.12	(108)	mao	17	0.07	(114)	bihg	11	0.04
(95)	hhaez	30	0.12	(109)	piypaxl	16	0.06	(114)	ehs	11	0.04
(95)	wehn	30	0.12	(109)	smaol	16	0.06	(114)	ayv	11	0.04
(96)	ao	29	0.11	(109)	seym	16	0.06	(115)	aen	10	0.04
(96)	baol	29	0.11	(109)	way	16	0.06	(115)	serkyuhlax	10	0.04
(97)	feys	28	0.11	(109)	gihv	16	0.06	(115)	nayt	10	0.04
(97)	wahnz	28	0.11	(110)	hhiaz	15	0.06	(115)	ayl	10	0.04
(97)	dihfraxnt	28	0.11	(110)	sheypt	15	0.06	(115)	lihtaxl	10	0.04
(98)	saynz	27	0.1	(110)	bihkohz	15	0.06	(115)	meybiy	10	0.04
(98)	wohz	27	0.1	(110)	bohksihz	15	0.06	(115)	smayliy	10	0.04
(98)	kahm	27	0.1	(110)	kaonaxz	15	0.06	(116)	fiyl	9	0.03
(99)	sahm	26	0.1	(110)	rihpiyt	15	0.06	(116)	taok	9	0.03
(99)	wohts	26	0.1	(111)	shuhd	14	0.05	(116)	hhaa	9	0.03
(100)	ihnsayd	25	0.1	(111)	tehl	14	0.05	(116)	wea	9	0.03
(100)	laast	25	0.1	(111)	thihngz	14	0.05	(116)	hhaendz	9	0.03
(100)	prihtih	25	0.1	(111)	duwihng	14	0.05	(116)	yuwv	9	0.03
(101)	ahp	24	0.09	(111)	sayn	14	0.05	(116)	dihd	9	0.03
(101)	yey	24	0.09	(112)	shao	13	0.05	(116)	dhowz	9	0.03
(101)	rowd	24	0.09	(112)	ferst	13	0.05	(116)	wihndow	9	0.03
(102)	aem	23	0.09	(112)	rowl	13	0.05	(116)	dey	9	0.03
(102)	wiyv	23	0.09	(112)	dhow	13	0.05	(116)	mihdaxl	9	0.03
(102)	geht	23	0.09	(112)	sahn	13	0.05	(116)	mawntihn	9	0.03
(102)	wayt	23	0.09	(112)	nehvax	13	0.05	(116)	leht	9	0.03
(103)	staat	22	0.09	(112)	laykt	13	0.05	(116)	iyvaxn	9	0.03
(103)	hhaepih	22	0.09	(112)	yaa	13	0.05	(116)	behtax	9	0.03

Table B.7 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(116)	iykwaxl	9	0.03	(119)	owvax	6	0.02	(121)	tehlihng	4	0.02
(116)	stihl	9	0.03	(119)	meyks	6	0.02	(121)	iykwihlaetaxraxl	4	0.02
(116)	dahzaxnt	9	0.03	(119)	hhaevaxnt	6	0.02	(121)	kohs	4	0.02
(116)	uh	9	0.03	(120)	yuwst	5	0.02	(121)	nialih	4	0.02
(117)	hhay	8	0.03	(120)	mahsaxnt	5	0.02	(121)	bihgihnihng	4	0.02
(117)	niyd	8	0.03	(120)	hhaevihng	5	0.02	(121)	awtsayd	4	0.02
(117)	sahmthihng	8	0.03	(120)	taokt	5	0.02	(121)	ihnstruhmaxnt	4	0.02
(117)	bohdihz	8	0.03	(120)	mowmaxnt	5	0.02	(121)	streyt	4	0.02
(117)	blohk	8	0.03	(120)	shaynz	5	0.02	(121)	ehnih	4	0.02
(117)	klaym	8	0.03	(120)	taxgehdhax	5	0.02	(121)	thruw	4	0.02
(117)	hhaed	8	0.03	(120)	hhihl	5	0.02	(121)	faom	4	0.02
(117)	saed	8	0.03	(120)	waw	5	0.02	(121)	wownt	4	0.02
(117)	hhowldihng	8	0.03	(120)	ehvrihwea	5	0.02	(121)	aahhaa	4	0.02
(117)	chohkaxlaxt	8	0.03	(120)	flaeg	5	0.02	(121)	stiyp	4	0.02
(118)	aodaxz	7	0.03	(120)	wey	5	0.02	(121)	geymz	4	0.02
(118)	luhkt	7	0.03	(120)	ehjhihz	5	0.02	(121)	stohp	4	0.02
(118)	saot	7	0.03	(120)	sahmtaymz	5	0.02	(121)	pleys	4	0.02
(118)	dheaz	7	0.03	(120)	iych	5	0.02	(121)	neym	4	0.02
(118)	bihld	7	0.03	(120)	klea	5	0.02	(121)	paat	4	0.02
(118)	rowboht	7	0.03	(120)	kahmz	5	0.02	(121)	biyts	4	0.02
(118)	wohntihd	7	0.03	(120)	dia	5	0.02	(121)	ihgzaektlih	4	0.02
(118)	kiyp	7	0.03	(120)	lernihd	5	0.02	(121)	aaftax	4	0.02
(118)	pliyz	7	0.03	(120)	sayd	5	0.02	(121)	boy	4	0.02
(118)	wihdhihn	7	0.03	(120)	pahmps	5	0.02	(121)	owkey	4	0.02
(118)	mayt	7	0.03	(120)	ehniymao	5	0.02	(121)	yaoz	4	0.02
(118)	wahndax	7	0.03	(120)	staend	5	0.02	(121)	sahnih	4	0.02
(118)	hhm	7	0.03	(120)	kiyps	5	0.02	(121)	beysihk	4	0.02
(118)	laynz	7	0.03	(120)	mawdh	5	0.02	(121)	towld	4	0.02
(118)	kiyn	7	0.03	(120)	aolweyz	5	0.02	(121)	maynd	4	0.02
(118)	axwey	7	0.03	(120)	bay	5	0.02	(121)	ohlrayt	4	0.02
(119)	ihntraxstihng	6	0.02	(120)	sihmbaxl	5	0.02	(121)	tohblaxrown	4	0.02
(119)	baod	6	0.02	(120)	triy	5	0.02	(121)	axlohng	4	0.02
(119)	bohtaxm	6	0.02	(120)	faynd	5	0.02	(122)	ehvrih	3	0.01
(119)	meyd	6	0.02	(120)	mahst	5	0.02	(122)	kwaotaxz	3	0.01
(119)	pleyd	6	0.02	(120)	frohm	5	0.02	(122)	aaaa	3	0.01
(119)	flaegz	6	0.02	(120)	yuwd	5	0.02	(122)	ehls	3	0.01
(119)	showz	6	0.02	(120)	bawns	5	0.02	(122)	pihlowz	3	0.01
(119)	pahmp	6	0.02	(120)	trayaenggyuhlax	5	0.02	(122)	taagiht	3	0.01
(119)	lernihng	6	0.02	(120)	aolrayt	5	0.02	(122)	kahlaxd	3	0.01
(119)	ahnfaochuhnaxtlih	6	0.02	(120)	wer	5	0.02	(122)	poynts	3	0.01
(119)	aydhax	6	0.02	(121)	gihvihng	4	0.02	(122)	fahnih	3	0.01
(119)	dahz	6	0.02	(121)	poynt	4	0.02	(122)	frehnd	3	0.01
(119)	taokihng	6	0.02	(121)	bihgax	4	0.02	(122)	klaymihng	3	0.01
(119)	smaylihng	6	0.02	(121)	axkrohs	4	0.02	(122)	aebsaxluwtlih	3	0.01
(119)	ihmpaotaxnt	6	0.02	(121)	blahd	4	0.02	(122)	hhyuwmaxn	3	0.01
(119)	meyk	6	0.02	(121)	ehnd	4	0.02	(122)	ownlih	3	0.01

Table B.7 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(122)	axlayv	3	0.01	(123)	seyihng	2	0.01	(123)	iyt	2	0.01
(122)	bawnsihng	3	0.01	(123)	ts	2	0.01	(123)	shaedowz	2	0.01
(122)	kyuwbz	3	0.01	(123)	sertaxnlih	2	0.01	(123)	bihhheyv	2	0.01
(122)	smaolax	3	0.01	(123)	leytax	2	0.01	(123)	ehnihthihng	2	0.01
(122)	fyuw	3	0.01	(123)	rehkaxgnayz	2	0.01	(123)	ohf	2	0.01
(122)	aolrehdih	3	0.01	(123)	kwaotax	2	0.01	(123)	ea	2	0.01
(122)	rihmehmbaxrihng	3	0.01	(123)	axgow	2	0.01	(123)	kahmihng	2	0.01
(122)	klawd	3	0.01	(123)	liyv	2	0.01	(123)	showihng	2	0.01
(122)	werd	3	0.01	(123)	sihks	2	0.01	(123)	prehzaxnts	2	0.01
(122)	hhaezaxnt	3	0.01	(123)	hhawehvax	2	0.01	(123)	pihraxmihdz	2	0.01
(122)	fahn	3	0.01	(123)	krehsaxnts	2	0.01	(123)	dhaen	2	0.01
(122)	ihnjhoyd	3	0.01	(123)	kaech	2	0.01	(123)	pleyst	2	0.01
(122)	glahm	3	0.01	(123)	fihnggax	2	0.01	(123)	fuhl	2	0.01
(122)	throw	3	0.01	(123)	faynaxl	2	0.01	(123)	frahnt	2	0.01
(122)	skuwl	3	0.01	(123)	muwv	2	0.01	(123)	rihmehmbaxd	2	0.01
(122)	wehlehntayns	3	0.01	(123)	tayps	2	0.01	(123)	slaytlih	2	0.01
(122)	sehkaxnd	3	0.01	(123)	mahndih	2	0.01	(123)	pihk	2	0.01
(122)	siyn	3	0.01	(123)	mehnih	2	0.01	(123)	wearaez	2	0.01
(122)	hhert	3	0.01	(123)	toyz	2	0.01	(123)	laajhax	2	0.01
(122)	rihmm	3	0.01	(123)	rihvayz	2	0.01	(123)	fihnihsht	2	0.01
(122)	fuhtbaol	3	0.01	(123)	kuhshaxn	2	0.01	(123)	kahvaxd	2	0.01
(122)	smuwdh	3	0.01	(123)	geyv	2	0.01	(123)	ohhh	2	0.01
(122)	dihng	3	0.01	(123)	geym	2	0.01	(123)	ahhhah	2	0.01
(122)	ihnjhoy	3	0.01	(123)	waonihngz	2	0.01	(123)	kahlax	2	0.01
(122)	miyn	3	0.01	(123)	pehnsaxl	2	0.01	(123)	axfreyd	2	0.01
(122)	hhey	3	0.01	(123)	iy	2	0.01	(123)	baolz	2	0.01
(122)	drohpt	3	0.01	(123)	grawnd	2	0.01	(123)	faastax	2	0.01
(122)	yuws	3	0.01	(123)	kahpaxl	2	0.01	(123)	ihnstehd	2	0.01
(122)	kaol	3	0.01	(123)	kerv	2	0.01	(123)	klows	2	0.01
(122)	thertih	3	0.01	(123)	sehshaxnz	2	0.01	(123)	faol	2	0.01
(122)	maethaxmaetihks	3	0.01	(123)	daats	2	0.01	(123)	wahns	2	0.01
(122)	mowst	3	0.01	(123)	meykihng	2	0.01	(123)	spiyd	2	0.01
(122)	uwps	3	0.01	(123)	hhmax	2	0.01	(123)	erlia	2	0.01
(122)	pihkchax	3	0.01	(123)	kaenaxt	2	0.01	(123)	laajh	2	0.01
(123)	gehtihng	2	0.01	(123)	biyihng	2	0.01	(123)	kaoz	2	0.01
(123)	paxhhaeps	2	0.01	(123)	taymz	2	0.01	(123)	drohp	2	0.01
(123)	trayd	2	0.01	(123)	kihk	2	0.01	(123)	deynjhax	2	0.01
(123)	ehndihng	2	0.01	(123)	siyihng	2	0.01	(123)	ehm	2	0.01
(123)	drohpihng	2	0.01	(123)	yuwzhaxlih	2	0.01	(123)	wiyk	2	0.01
(123)	keafuhl	2	0.01	(123)	uhps	2	0.01	(123)	thihng	2	0.01
(123)	faxrehvax	2	0.01	(123)	jhoyn	2	0.01	(123)	nyuw	2	0.01
(123)	siymd	2	0.01	(123)	jhohb	2	0.01	(123)	chehst	2	0.01
(123)	throwihng	2	0.01	(123)	fihftih	2	0.01	(123)	sahmwahn	2	0.01
(123)	rohng	2	0.01	(123)	werk	2	0.01	(124)	mihniht	1	0
(123)	showd	2	0.01	(123)	brihliant	2	0.01	(124)	nahthihng	1	0
(123)	eyay	2	0.01	(123)	kea	2	0.01	(124)	sh	1	0

Table B.7 – *Continued from previous page*

rank word	cnt	%	*rank* word	cnt	%	*rank* word	cnt	%
(124) riych	1	0	(124) ehndlihs	1	0	(124) prohbaxblih	1	0
(124) kwihklih	1	0	(124) sheyk	1	0	(124) paxtihkyuhlax	1	0
(124) cheynjhd	1	0	(124) aestraxnaot	1	0	(124) ehvrihbohdih	1	0
(124) bihkahmz	1	0	(124) buwlzay	1	0	(124) bowth	1	0
(124) dhearax	1	0	(124) bayt	1	0	(124) hhahnggrih	1	0
(124) kaxnfyuwzd	1	0	(124) kaot	1	0	(124) aeksihdaxnt	1	0
(124) krohp	1	0	(124) mohrnihn	1	0	(124) aam	1	0
(124) hhhh	1	0	(124) lay	1	0	(124) tiy	1	0
(124) klawdz	1	0	(124) sertaxn	1	0	(124) meyz	1	0
(124) ehf	1	0	(124) shaynihng	1	0	(124) priy	1	0
(124) hhuwps	1	0	(124) wow	1	0	(124) ahdhaxz	1	0
(124) weyv	1	0	(124) gohn	1	0	(124) pehtaxlz	1	0
(124) shuhdaxnt	1	0	(124) ehvrihthihng	1	0	(124) ehks	1	0
(124) berthdey	1	0	(124) sk	1	0	(124) prohpax	1	0
(124) wohtehvax	1	0	(124) grihp	1	0	(124) klowsax	1	0
(124) priht	1	0	(124) maonihng	1	0	(124) rehd	1	0
(124) flaet	1	0	(124) s	1	0	(124) prohblaxm	1	0
(124) behst	1	0	(124) layts	1	0	(124) sahch	1	0
(124) seh	1	0	(124) row	1	0	(124) wuht	1	0
(124) mey	1	0	(124) hheyts	1	0	(124) yiaz	1	0
(124) muwns	1	0	(124) seynt	1	0	(124) riychihng	1	0
(124) kaxrehkt	1	0	(124) piy	1	0	(124) perfihkt	1	0
(124) kertaxnz	1	0	(124) faynaxlih	1	0	(124) mayn	1	0
(124) gerlfrehnd	1	0	(124) ohntax	1	0	(124) spiykihng	1	0
(124) axl	1	0	(124) staendihng	1	0	(124) ruwmz	1	0
(124) sao	1	0	(124) staa	1	0	(124) fawnd	1	0
(124) saytlih	1	0	(124) aaskt	1	0	(124) ohps	1	0
(124) mihsduw	1	0	(124) puhsh	1	0	(124) ruwm	1	0
(124) miytihng	1	0	(124) wuhdaxnt	1	0	(124) sehntax	1	0
(124) hhaxm	1	0	(124) ihmehmbax	1	0	(124) yiy	1	0
(124) thahm	1	0	(124) trayaeng	1	0	(124) axhhehd	1	0
(124) shuwt	1	0	(124) truw	1	0	(124) lehs	1	0
(124) nohnoh	1	0	(124) wayl	1	0	(124) kaxmpliytlih	1	0
(124) hhuw	1	0	(124) yaosehlf	1	0	(124) poyntiy	1	0
(124) breyk	1	0	(124) fihnihsh	1	0	(124) neymz	1	0
(124) rihmayndz	1	0	(124) ehndz	1	0	(124) feyvaxrihts	1	0
(124) sheh	1	0	(124) sehmih	1	0	(124) skayz	1	0
(124) striyt	1	0	(124) ahndaxstaend	1	0	(124) maxstaash	1	0
(124) kaadz	1	0	(124) fayn	1	0	(124) trax	1	0
(124) sihmbaxlz	1	0	(124) lohlihpohp	1	0	(124) feysihz	1	0
(124) maechihz	1	0	(124) hhaad	1	0	(124) baejhihz	1	0
(124) owldax	1	0	(124) hhawzihz	1	0	(124) nowtihst	1	0
(124) diy	1	0	(124) axntihl	1	0	(124) strohng	1	0
(124) slowp	1	0	(124) trayihng	1	0	(124) mehzhax	1	0
(124) kehpt	1	0	(124) guhdnaxs	1	0	(124) rowbohts	1	0
(124) frehndlih	1	0	(124) riycht	1	0	(124) lahvd	1	0

Table B.7 – *Continued from previous page*

rank word	cnt	%	*rank* word	cnt	%	*rank* word	cnt	%
(124) noyz	1	0	(124) iyjhihpt	1	0	(124) lax	1	0
(124) kr	1	0	(124) faoth	1	0	(124) meynlih	1	0
(124) ayd	1	0	(124) sihmehtrihkaxl	1	0	(124) fiylihng	1	0
(124) dhih	1	0	(124) dihfraxns	1	0	(124) iyzih	1	0
(124) luhkawt	1	0	(124) flao	1	0	(124) hhohraxbaxl	1	0
(124) streytaxwey	1	0	(124) daak	1	0	(124) ohfaxn	1	0
(124) perfihktlih	1	0	(124) sihmbohlayzihz	1	0	(124) yuwzihng	1	0
(124) ae	1	0	(124) staatihd	1	0	(124) aess	1	0
(124) hhohldihn	1	0	(124) weyz	1	0	(124) yeht	1	0
(124) faa	1	0	(124) jhoynd	1	0	(124) ihnax	1	0
(124) sahmtaym	1	0	(124) shaht	1	0	(124) faolihng	1	0
(124) axtehnshaxn	1	0	(124) eynshaxnt	1	0	(124) pihraxmihd	1	0
(124) drao	1	0	(124) daw	1	0	(124) low	1	0
(124) duwihn	1	0	(124) ahs	1	0	(124) tehmptihng	1	0
(124) kowld	1	0	(124) poyntihd	1	0	(124) biytihng	1	0
(124) kahvaxz	1	0	(124) miyt	1	0	(124) flawaxz	1	0
(124) aerowz	1	0	(124) ael	1	0	(124) hhaadlih	1	0
(124) sihtihng	1	0	(124) naydhax	1	0	(124) tawax	1	0
(124) mihrax	1	0	(124) bihts	1	0	(124) teybaxl	1	0
(124) ehvax	1	0	(124) sihmaxlax	1	0	(124) fihnggaxz	1	0
(124) tehnihs	1	0	(124) buwmaxraeng	1	0	(124) naomaxlih	1	0
(124) saxk	1	0	(124) klowslih	1	0	(124) aachaxrih	1	0
(124) saht	1	0	(124) mow	1	0	(124) fraengk	1	0
(124) hhae	1	0	(124) gerl	1	0	(124) awtax	1	0
(124) staendihn	1	0	(124) taxnayt	1	0	(124) sheykihng	1	0
(124) faast	1	0	(124) lihsaxn	1	0	(124) krehs	1	0
(124) shahflihng	1	0	(124) baed	1	0	(124) wehdhax	1	0
(124) sih	1	0	(124) axdmiht	1	0	(124) key	1	0
(124) rihng	1	0	(124) ihntax	1	0	(124) earia	1	0
(124) spehshaxl	1	0	(124) hhahzbaxnd	1	0	(124) hhehd	1	0
(124) saxpowz	1	0	(124) luwkihn	1	0	(124) dhahs	1	0
(124) lern	1	0	(124) greyt	1	0	(124) ihtsehlf	1	0
(124) serk	1	0	(124) faolz	1	0	(124) gohtax	1	0
(124) hhehdseht	1	0	(124) yuwr	1	0	(124) klowzd	1	0
(124) st	1	0	(124) brayt	1	0	(124) ihmprehst	1	0
(124) wayf	1	0	(124) miynihng	1	0	(124) wernt	1	0
(124) axfehkshaxn	1	0	(124) nayslih	1	0	(124) bahmp	1	0
(124) saots	1	0	(124) liyst	1	0	(124) fiyldz	1	0
(124) chohkaxlaxts	1	0	(124) trohnohmiy	1	0	(124) miytihngz	1	0
(124) kaxmpliyt	1	0	(124) baelaxnst	1	0	(124) stahk	1	0
(124) piysihz	1	0	(124) ohnihstlih	1	0			
(124) faomihng	1	0	(124) thaot	1	0			
(124) yax	1	0	(124) smayld	1	0			

Table B.8: ***Complete word-frequencies of salient words in prohibition experiment.*** *Listed are the rank, word count (*cnt*) and the percentage relative to the total number of words in the experiment across all participants and sessions.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	skwea	327	4.4	(39)	wehl	40	0.54	(58)	lahv	16	0.22
(2)	trayaenggaxl	302	4.06	(40)	kaant	39	0.52	(58)	blaek	16	0.22
(3)	serkaxl	285	3.83	(41)	gow	38	0.51	(59)	kwayt	15	0.2
(4)	now	272	3.66	(42)	dhea	37	0.5	(59)	smayl	15	0.2
(5)	wahn	250	3.36	(42)	naw	37	0.5	(59)	tray	15	0.2
(6)	hhaat	248	3.34	(43)	ow	34	0.46	(59)	ohn	15	0.2
(7)	dhihs	230	3.09	(43)	hhowld	34	0.46	(59)	thriy	15	0.2
(8)	muwn	208	2.8	(44)	tahch	32	0.43	(60)	puht	14	0.19
(9)	uhkey	171	2.3	(45)	dawn	31	0.42	(60)	ihnsayd	14	0.19
(10)	sheyp	159	2.14	(45)	noht	31	0.42	(60)	dhehn	14	0.19
(11)	yehs	155	2.08	(46)	dhaets	30	0.4	(60)	nehkst	14	0.19
(12)	krehsaxnt	149	2	(46)	ehksaxlaxnt	30	0.4	(60)	baorihng	14	0.19
(13)	layk	134	1.8	(47)	rialih	29	0.39	(61)	prihtih	13	0.17
(14)	serkaxlz	129	1.74	(47)	duw	29	0.39	(61)	woht	13	0.17
(15)	axgehn	112	1.51	(48)	saydz	28	0.38	(61)	goht	13	0.17
(16)	guhd	107	1.44	(48)	sohrih	28	0.38	(61)	hhaxlow	13	0.17
(17)	iht	101	1.36	(49)	feyvaxriht	27	0.36	(62)	feys	12	0.16
(18)	vehrih	93	1.25	(49)	siy	27	0.36	(63)	axlawd	11	0.15
(19)	uhey	88	1.18	(50)	aet	25	0.34	(63)	ay	11	0.15
(20)	diychiy	83	1.12	(51)	lehts	23	0.31	(63)	bohksihz	11	0.15
(21)	yuw	79	1.06	(52)	sow	22	0.3	(64)	gowihng	10	0.13
(22)	rawnd	76	1.02	(52)	kaen	22	0.3	(64)	sahn	10	0.13
(23)	pley	75	1.01	(53)	aem	21	0.28	(64)	hhaepih	10	0.13
(24)	bohks	72	0.97	(53)	ax	21	0.28	(64)	kyuwb	10	0.13
(25)	diyjhiy	67	0.9	(54)	tuw	20	0.27	(64)	dihfraxnt	10	0.13
(26)	dhaet	66	0.89	(54)	kaold	20	0.27	(64)	wayt	10	0.13
(26)	rayt	66	0.89	(54)	nays	20	0.27	(65)	sey	9	0.12
(27)	skweaz	62	0.83	(55)	aend	19	0.26	(65)	serkyuhlax	9	0.12
(27)	ihz	62	0.83	(55)	wohnt	19	0.26	(65)	prehzaxnt	9	0.12
(28)	dahn	60	0.81	(55)	baht	19	0.26	(65)	ihzaxnt	9	0.12
(29)	rihmehmbax	59	0.79	(55)	thihngk	19	0.26	(65)	wihndow	9	0.12
(30)	luhk	52	0.7	(55)	baol	19	0.26	(65)	smaol	9	0.12
(31)	taxdey	51	0.69	(55)	yey	19	0.26	(65)	know	9	0.12
(32)	trayaenggaxlz	50	0.67	(55)	teyk	19	0.26	(66)	lahvlih	8	0.11
(33)	axbawt	49	0.66	(56)	uwm	18	0.24	(66)	staat	8	0.11
(34)	aa	47	0.63	(56)	hhaev	18	0.24	(66)	ao	8	0.11
(35)	dhax	45	0.61	(56)	taym	18	0.24	(66)	waonihng	8	0.11
(35)	downt	45	0.61	(56)	fao	18	0.24	(66)	piypaxl	8	0.11
(35)	yea	45	0.61	(57)	baek	17	0.23	(66)	nehvax	8	0.11
(36)	axnahdhax	44	0.59	(57)	ihts	17	0.23	(66)	aolsow	8	0.11
(37)	hhaats	43	0.58	(57)	wiy	17	0.23	(66)	seym	8	0.11
(38)	hhia	41	0.55	(58)	saynz	16	0.22	(66)	biht	8	0.11
(39)	sheyps	40	0.54	(58)	aol	16	0.22	(66)	way	8	0.11

Table B.8 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(67)	aodaxz	7	0.09	(69)	klaym	5	0.07	(71)	ahp	3	0.04
(67)	bihfao	7	0.09	(69)	behtax	5	0.07	(71)	wia	3	0.04
(67)	rowl	7	0.09	(69)	iykwaxl	5	0.07	(71)	aebsaxluwtlih	3	0.04
(67)	laast	7	0.09	(69)	axrawnd	5	0.07	(71)	hhay	3	0.04
(67)	dey	7	0.09	(69)	dhiyz	5	0.07	(71)	hhaend	3	0.04
(67)	wihdh	7	0.09	(69)	smaylihng	5	0.07	(71)	flaegz	3	0.04
(67)	mihdaxl	7	0.09	(69)	kahm	5	0.07	(71)	nialih	3	0.04
(67)	lohts	7	0.09	(69)	sehd	5	0.07	(71)	bihgihnihng	3	0.04
(67)	geht	7	0.09	(69)	mao	5	0.07	(71)	wey	3	0.04
(67)	luhks	7	0.09	(69)	dihdaxnt	5	0.07	(71)	bawnsihng	3	0.04
(67)	bihldihng	7	0.09	(69)	wohts	5	0.07	(71)	pahmp	3	0.04
(67)	hhm	7	0.09	(70)	aolrayt	4	0.05	(71)	ihnstruhmaxnt	3	0.04
(67)	shaap	7	0.09	(70)	fiyl	4	0.05	(71)	streyt	3	0.04
(67)	kaonaxz	7	0.09	(70)	dhehm	4	0.05	(71)	tohp	3	0.04
(67)	sayn	7	0.09	(70)	axkrohs	4	0.05	(71)	blohk	3	0.04
(67)	chohkaxlaxt	7	0.09	(70)	wea	4	0.05	(71)	lernihng	3	0.04
(68)	nayt	6	0.08	(70)	waw	4	0.05	(71)	wihdhihn	3	0.04
(68)	hhaa	6	0.08	(70)	flaeg	4	0.05	(71)	klea	3	0.04
(68)	shao	6	0.08	(70)	sahmtaymz	4	0.05	(71)	aahhaa	3	0.04
(68)	ferst	6	0.08	(70)	mawntihn	4	0.05	(71)	hhaw	3	0.04
(68)	ihntraxstihd	6	0.08	(70)	wohntihd	4	0.05	(71)	lernihd	3	0.04
(68)	aekchualih	6	0.08	(70)	pliyz	4	0.05	(71)	ohv	3	0.04
(68)	meybiy	6	0.08	(70)	miy	4	0.05	(71)	stohp	3	0.04
(68)	bohdihz	6	0.08	(70)	wahndax	4	0.05	(71)	skuwl	3	0.04
(68)	smayliy	6	0.08	(70)	klehvax	4	0.05	(71)	ehniymao	3	0.04
(68)	miynz	6	0.08	(70)	ehs	4	0.05	(71)	gowz	3	0.04
(68)	yaa	6	0.08	(70)	dhey	4	0.05	(71)	neym	3	0.04
(68)	wahnz	6	0.08	(70)	ihf	4	0.05	(71)	wehn	3	0.04
(68)	thaengk	6	0.08	(70)	stihl	4	0.05	(71)	sehkaxnd	3	0.04
(68)	skay	6	0.08	(70)	ihgzaektlih	4	0.05	(71)	bihg	3	0.04
(68)	duwihng	6	0.08	(70)	beysihk	4	0.05	(71)	saed	3	0.04
(68)	luhkihng	6	0.08	(70)	kiyn	4	0.05	(71)	ahdhax	3	0.04
(68)	wihch	6	0.08	(70)	ohlrayt	4	0.05	(71)	hhert	3	0.04
(68)	mahch	6	0.08	(71)	thertih	3	0.04	(71)	er	3	0.04
(68)	loht	6	0.08	(71)	kwaotaxz	3	0.04	(71)	pleyihng	3	0.04
(68)	laynz	6	0.08	(71)	ihntraxstihng	3	0.04	(71)	aolweyz	3	0.04
(68)	uh	6	0.08	(71)	aaaa	3	0.04	(71)	owkey	3	0.04
(69)	baod	5	0.07	(71)	yao	3	0.04	(71)	sihmbaxl	3	0.04
(69)	hhaaf	5	0.07	(71)	luhkt	3	0.04	(71)	triy	3	0.04
(69)	hhaendz	5	0.07	(71)	bohtaxm	3	0.04	(71)	aant	3	0.04
(69)	dhow	5	0.07	(71)	pihlowz	3	0.04	(71)	ihmpaotaxnt	3	0.04
(69)	showz	5	0.07	(71)	tehl	3	0.04	(71)	hhey	3	0.04
(69)	laykt	5	0.07	(71)	taagiht	3	0.04	(71)	rowd	3	0.04
(69)	rowboht	5	0.07	(71)	wohnax	3	0.04	(71)	rihpiyt	3	0.04
(69)	ahnfaochuhnaxtlih	5	0.07	(71)	poynts	3	0.04	(71)	tohblaxrown	3	0.04
(69)	blohks	5	0.07	(71)	trayaenggyuhlax	3	0.04	(71)	axwey	3	0.04

Table B.8 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(71)	uwps	3	0.04	(72)	sahmthihng	2	0.03	(72)	erlia	2	0.03
(71)	pihkchax	3	0.04	(72)	sheypt	2	0.03	(72)	laajh	2	0.03
(72)	ayv	2	0.03	(72)	bihkohz	2	0.03	(72)	owvax	2	0.03
(72)	thihngz	2	0.03	(72)	hhmax	2	0.03	(72)	dahzaxnt	2	0.03
(72)	gehtihng	2	0.03	(72)	kaenaxt	2	0.03	(72)	drohpt	2	0.03
(72)	paxhhaeps	2	0.03	(72)	ehjhihz	2	0.03	(72)	maethaxmaetihks	2	0.03
(72)	wuhd	2	0.03	(72)	uhps	2	0.03	(72)	hhaevaxnt	2	0.03
(72)	drohpihng	2	0.03	(72)	sahm	2	0.03	(72)	chehst	2	0.03
(72)	keafuhl	2	0.03	(72)	iych	2	0.03	(72)	axlohng	2	0.03
(72)	mahsaxnt	2	0.03	(72)	fihftih	2	0.03	(73)	kwihklih	1	0.01
(72)	poynt	2	0.03	(72)	brihliant	2	0.03	(73)	ehvrih	1	0.01
(72)	taokt	2	0.03	(72)	rihmehmbaxrihng	2	0.03	(73)	shuhd	1	0.01
(72)	mowmaxnt	2	0.03	(72)	shaedowz	2	0.03	(73)	faxrehvax	1	0.01
(72)	bihgax	2	0.03	(72)	fahn	2	0.03	(73)	siymd	1	0.01
(72)	hhaez	2	0.03	(72)	gohnax	2	0.03	(73)	klawdz	1	0.01
(72)	sertaxnlih	2	0.03	(72)	bihhheyv	2	0.03	(73)	hhuwps	1	0.01
(72)	shaynz	2	0.03	(72)	ihnjhoyd	2	0.03	(73)	weyv	1	0.01
(72)	leytax	2	0.03	(72)	mayt	2	0.03	(73)	berthdey	1	0.01
(72)	rehkaxgnayz	2	0.03	(72)	iyvaxn	2	0.03	(73)	wohtehvax	1	0.01
(72)	aym	2	0.03	(72)	dia	2	0.03	(73)	flaet	1	0.01
(72)	kwaotax	2	0.03	(72)	aydhax	2	0.03	(73)	behst	1	0.01
(72)	yua	2	0.03	(72)	prehzaxnts	2	0.03	(73)	kaxrehkt	1	0.01
(72)	hhawehvax	2	0.03	(72)	hhaed	2	0.03	(73)	kertaxnz	1	0.01
(72)	krehsaxnts	2	0.03	(72)	throw	2	0.03	(73)	gerlfrehnd	1	0.01
(72)	wohz	2	0.03	(72)	wehlehntayns	2	0.03	(73)	axl	1	0.01
(72)	faynaxl	2	0.03	(72)	fuhl	2	0.03	(73)	hhaevihng	1	0.01
(72)	saot	2	0.03	(72)	rihmehmbaxd	2	0.03	(73)	saytlih	1	0.01
(72)	blahd	2	0.03	(72)	biyts	2	0.03	(73)	mihsduw	1	0.01
(72)	lihtaxl	2	0.03	(72)	hhowldihng	2	0.03	(73)	eyay	1	0.01
(72)	mahndih	2	0.03	(72)	rihmm	2	0.03	(73)	miytihng	1	0.01
(72)	toyz	2	0.03	(72)	fuhtbaol	2	0.03	(73)	hhaxm	1	0.01
(72)	ehvrihwea	2	0.03	(72)	axfreyd	2	0.03	(73)	ts	1	0.01
(72)	kuhshaxn	2	0.03	(72)	kahlax	2	0.03	(73)	thahm	1	0.01
(72)	waonihngz	2	0.03	(72)	baolz	2	0.03	(73)	nohnoh	1	0.01
(72)	pehnsaxl	2	0.03	(72)	ihnstehd	2	0.03	(73)	shael	1	0.01
(72)	kerv	2	0.03	(72)	yaoz	2	0.03	(73)	wiyv	1	0.01
(72)	sehshaxnz	2	0.03	(72)	show	2	0.03	(73)	striyt	1	0.01
(72)	daats	2	0.03	(72)	smuwdh	2	0.03	(73)	kaadz	1	0.01
(72)	niyd	2	0.03	(72)	spiyd	2	0.03	(73)	maechihz	1	0.01
(72)	awtsayd	2	0.03	(72)	towld	2	0.03	(73)	sihks	1	0.01
(72)	hhiaz	2	0.03	(72)	ihnjhoy	2	0.03	(73)	slowp	1	0.01
(73)	kahlaxd	1	0.01	(73)	lohlihpohp	1	0.01	(73)	noyz	1	0.01
(73)	kaech	1	0.01	(73)	awt	1	0.01	(73)	streytaxwey	1	0.01
(73)	frehndlih	1	0.01	(73)	hhawzihz	1	0.01	(73)	perfihktlih	1	0.01
(73)	taxgehdhax	1	0.01	(73)	axntihl	1	0.01	(73)	hhohldihn	1	0.01
(73)	fihnggax	1	0.01	(73)	guhdnaxs	1	0.01	(73)	ea	1	0.01

Table B.8 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(73)	ehndlihs	1	0.01	(73)	paxtihkyuhlax	1	0.01	(73)	sahmtaym	1	0.01
(73)	aestraxnaot	1	0.01	(73)	kyuwbz	1	0.01	(73)	axtehnshaxn	1	0.01
(73)	frehnd	1	0.01	(73)	ehvrihbohdih	1	0.01	(73)	duwihn	1	0.01
(73)	klaymihng	1	0.01	(73)	ehnih	1	0.01	(73)	showihng	1	0.01
(73)	buwlzay	1	0.01	(73)	taymz	1	0.01	(73)	stiyp	1	0.01
(73)	ehnd	1	0.01	(73)	aeksihdaxnt	1	0.01	(73)	glahm	1	0.01
(73)	pleyd	1	0.01	(73)	meyz	1	0.01	(73)	aerowz	1	0.01
(73)	tehlihng	1	0.01	(73)	bihld	1	0.01	(73)	pihraxmihdz	1	0.01
(73)	mohrnihn	1	0.01	(73)	ahdhaxz	1	0.01	(73)	geymz	1	0.01
(73)	hhyuwmaxn	1	0.01	(73)	pehtaxlz	1	0.01	(73)	saht	1	0.01
(73)	tayps	1	0.01	(73)	klowsax	1	0.01	(73)	pahmps	1	0.01
(73)	sertaxn	1	0.01	(73)	rehd	1	0.01	(73)	faast	1	0.01
(73)	shaynihng	1	0.01	(73)	prohblaxm	1	0.01	(73)	shahflihng	1	0.01
(73)	hhihl	1	0.01	(73)	yiaz	1	0.01	(73)	rihng	1	0.01
(73)	ehvrihthihng	1	0.01	(73)	riychihng	1	0.01	(73)	spehshaxl	1	0.01
(73)	grihp	1	0.01	(73)	yuwzhaxlih	1	0.01	(73)	pleys	1	0.01
(73)	dheaz	1	0.01	(73)	fyuw	1	0.01	(73)	serk	1	0.01
(73)	maonihng	1	0.01	(73)	ruwmz	1	0.01	(73)	saxpowz	1	0.01
(73)	layts	1	0.01	(73)	faom	1	0.01	(73)	hhehdseht	1	0.01
(73)	hheyts	1	0.01	(73)	ohps	1	0.01	(73)	pleyst	1	0.01
(73)	piy	1	0.01	(73)	wownt	1	0.01	(73)	wayf	1	0.01
(73)	faynaxlih	1	0.01	(73)	kiyp	1	0.01	(73)	axfehkshaxn	1	0.01
(73)	rihvayz	1	0.01	(73)	kaxmpliytlih	1	0.01	(73)	saots	1	0.01
(73)	dihd	1	0.01	(73)	kea	1	0.01	(73)	chohkaxlaxts	1	0.01
(73)	staendihng	1	0.01	(73)	poyntiy	1	0.01	(73)	frahnt	1	0.01
(73)	geyv	1	0.01	(73)	aolrehdih	1	0.01	(73)	kaxmpliyt	1	0.01
(73)	axlayv	1	0.01	(73)	leht	1	0.01	(73)	staend	1	0.01
(73)	staa	1	0.01	(73)	feyvaxrihts	1	0.01	(73)	paat	1	0.01
(73)	geym	1	0.01	(73)	klawd	1	0.01	(73)	kiyps	1	0.01
(73)	iykwihlaetaxraxl	1	0.01	(73)	skayz	1	0.01	(73)	iyjhihpt	1	0.01
(73)	truw	1	0.01	(73)	hhaezaxnt	1	0.01	(73)	sihmehtrihkaxl	1	0.01
(73)	iy	1	0.01	(73)	feysihz	1	0.01	(73)	dihfraxns	1	0.01
(73)	grawnd	1	0.01	(73)	baejhihz	1	0.01	(73)	dahz	1	0.01
(73)	wayl	1	0.01	(73)	nowtihst	1	0.01	(73)	sihmbohlayzihz	1	0.01
(73)	ahndaxstaend	1	0.01	(73)	rowbohts	1	0.01	(73)	staatihd	1	0.01
(73)	fayn	1	0.01	(73)	mehzhax	1	0.01	(73)	jhoynd	1	0.01
(73)	dhowz	1	0.01	(73)	kr	1	0.01	(73)	eynshaxnt	1	0.01
(73)	slaytlih	1	0.01	(73)	luwkihn	1	0.01	(73)	biytihng	1	0.01
(73)	poyntihd	1	0.01	(73)	brayt	1	0.01	(73)	flawaxz	1	0.01
(73)	ael	1	0.01	(73)	siym	1	0.01	(73)	drohp	1	0.01
(73)	naydhax	1	0.01	(73)	trohnohmiy	1	0.01	(73)	teybaxl	1	0.01
(73)	awax	1	0.01	(73)	klows	1	0.01	(73)	deynjhax	1	0.01
(73)	bihts	1	0.01	(73)	baelaxnst	1	0.01	(73)	yuws	1	0.01
(73)	wearaez	1	0.01	(73)	wahns	1	0.01	(73)	aachaxrih	1	0.01
(73)	laajhax	1	0.01	(73)	faynd	1	0.01	(73)	wiyk	1	0.01
(73)	buwmaxraeng	1	0.01	(73)	meynlih	1	0.01	(73)	sheykihng	1	0.01

Table B.8 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(73)	klowslih	1	0.01	(73)	iyzih	1	0.01	(73)	krehs	1	0.01
(73)	gerl	1	0.01	(73)	ohfaxn	1	0.01	(73)	key	1	0.01
(73)	taokihng	1	0.01	(73)	yeht	1	0.01	(73)	earia	1	0.01
(73)	lihsaxn	1	0.01	(73)	aess	1	0.01	(73)	hhehd	1	0.01
(73)	boy	1	0.01	(73)	faolihng	1	0.01	(73)	klowzd	1	0.01
(73)	axdmiht	1	0.01	(73)	pihraxmihd	1	0.01	(73)	ihmprehst	1	0.01
(73)	hhahzbaxnd	1	0.01	(73)	may	1	0.01	(73)	wernt	1	0.01
(73)	ohhh	1	0.01	(73)	tehmptihng	1	0.01	(73)	fiyldz	1	0.01

Table B.9: ***Complete word-frequencies of all words in the experiment of Saunders et al. (2012)****. Listed are the rank, word count (*cnt*) and the percentage relative to the total number of words in the experiment across all participants and sessions.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	ax	702	8.71	(30)	layk	65	0.81	(49)	dhehn	29	0.36
(2)	dhihs	367	4.55	(31)	staa	61	0.76	(50)	aen	28	0.35
(3)	bluw	347	4.31	(32)	woht	56	0.69	(50)	waots	28	0.35
(4)	ihz	322	4	(33)	sahn	54	0.67	(50)	smayl	28	0.35
(5)	aend	314	3.9	(34)	uhkey	50	0.62	(50)	now	28	0.35
(6)	rehd	302	3.75	(34)	naw	50	0.62	(50)	dhea	28	0.35
(7)	griyn	286	3.55	(35)	wihdh	48	0.6	(51)	hhaez	26	0.32
(8)	dhax	265	3.29	(36)	sow	46	0.57	(52)	sheyps	25	0.31
(9)	dhaets	237	2.94	(37)	yea	44	0.55	(52)	kaen	25	0.31
(10)	yuw	194	2.41	(38)	bihgax	41	0.51	(53)	ihm	24	0.3
(11)	ihts	161	2	(38)	wayt	41	0.51	(54)	show	23	0.29
(12)	hhaat	160	1.99	(39)	laajh	40	0.5	(55)	bihg	22	0.27
(13)	serkaxl	149	1.85	(40)	aet	39	0.48	(55)	dahn	22	0.27
(14)	aerow	148	1.84	(41)	kahlax	37	0.46	(55)	vehrih	22	0.27
(15)	sayd	146	1.81	(42)	aa	36	0.45	(56)	downt	21	0.26
(16)	krohs	120	1.49	(42)	ihn	36	0.45	(56)	er	20	0.25
(17)	hhia	112	1.39	(43)	siy	35	0.43	(57)	poyntihng	19	0.24
(18)	wiy	111	1.38	(43)	baodax	35	0.43	(57)	smaolax	19	0.24
(19)	ohn	110	1.37	(43)	baekgrawnd	35	0.43	(58)	tray	17	0.21
(20)	muwn	95	1.18	(43)	tuw	35	0.43	(58)	krehsaxnt	17	0.21
(21)	wahn	94	1.17	(44)	axbawt	34	0.42	(58)	gow	17	0.21
(22)	dhaet	89	1.1	(44)	guhd	34	0.42	(58)	hhaxlow	17	0.21
(23)	sheyp	88	1.09	(45)	diychiy	33	0.41	(59)	hhaw	16	0.2
(24)	rayt	87	1.08	(45)	yehs	33	0.41	(59)	ihf	16	0.2
(24)	bohks	87	1.08	(46)	luhk	32	0.4	(59)	baht	16	0.2
(25)	hhaev	80	0.99	(46)	miydiam	32	0.4	(59)	saydz	16	0.2
(26)	goht	79	0.98	(47)	sahm	31	0.38	(59)	know	16	0.2
(27)	smaol	76	0.94	(48)	wehl	30	0.37	(60)	brihliant	15	0.19
(28)	iht	75	0.93	(48)	duw	30	0.37	(60)	showihng	15	0.19
(29)	skwea	69	0.86	(49)	ay	29	0.36	(60)	kaxrehkt	15	0.19

Table B.9 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(60)	uwm	15	0.19	(69)	nays	6	0.07	(73)	baek	2	0.02
(60)	axgehn	15	0.19	(70)	dhehm	5	0.06	(73)	mayt	2	0.02
(60)	ahpwaxdz	15	0.19	(70)	kyuwb	5	0.06	(73)	baod	2	0.02
(61)	tern	14	0.17	(70)	behtax	5	0.06	(73)	ehls	2	0.02
(61)	gowihng	14	0.17	(70)	yp	5	0.06	(73)	skay	2	0.02
(61)	gohnax	14	0.17	(70)	seyihng	5	0.06	(73)	duwihng	2	0.02
(61)	luhkihng	14	0.17	(70)	fao	5	0.06	(73)	ihl	2	0.02
(61)	lehts	14	0.17	(70)	bihfao	5	0.06	(73)	dawn	2	0.02
(61)	wiyv	14	0.17	(70)	smaolihst	5	0.06	(73)	sihtihng	2	0.02
(61)	ahdhax	14	0.17	(70)	laast	5	0.06	(73)	dhaen	2	0.02
(61)	noht	14	0.17	(70)	showz	5	0.06	(73)	ayl	2	0.02
(61)	mihdaxl	14	0.17	(70)	saxrawndihd	5	0.06	(73)	sihks	2	0.02
(62)	taxdey	13	0.16	(71)	yao	4	0.05	(73)	sohrih	2	0.02
(62)	thihngk	13	0.16	(71)	kaent	4	0.05	(73)	wia	2	0.02
(63)	shael	12	0.15	(71)	jhahst	4	0.05	(73)	ferst	2	0.02
(64)	dihfraxnt	11	0.14	(71)	wihch	4	0.05	(73)	nehkst	2	0.02
(64)	kahlaxz	11	0.14	(71)	feysihng	4	0.05	(73)	dihd	2	0.02
(64)	rawnd	11	0.14	(71)	dhiyz	4	0.05	(73)	sahmthihng	2	0.02
(64)	ihznt	11	0.14	(71)	laajhax	4	0.05	(73)	feys	2	0.02
(65)	aym	10	0.12	(71)	wihl	4	0.05	(73)	mao	2	0.02
(65)	axnahdhax	10	0.12	(71)	sfia	4	0.05	(73)	laykt	2	0.02
(66)	aol	9	0.11	(71)	owkey	4	0.05	(73)	sahmtaymz	2	0.02
(66)	wohnt	9	0.11	(71)	faynd	4	0.05	(74)	aolrehdih	1	0.01
(66)	ao	9	0.11	(71)	wey	4	0.05	(74)	wohchihng	1	0.01
(66)	aolsow	9	0.11	(71)	owvax	4	0.05	(74)	sey	1	0.01
(67)	wahnz	8	0.1	(72)	dihdnt	3	0.04	(74)	shuhd	1	0.01
(67)	bohksihz	8	0.1	(72)	geht	3	0.04	(74)	feysihz	1	0.01
(67)	ohv	8	0.1	(72)	ihgzaampaxl	3	0.04	(74)	lahvlih	1	0.01
(67)	ihnsayd	8	0.1	(72)	feyvaxriht	3	0.04	(74)	taok	1	0.01
(67)	sayzd	8	0.1	(72)	shao	3	0.04	(74)	dhiy	1	0.01
(67)	aez	8	0.1	(72)	paetaxnz	3	0.04	(74)	rohng	1	0.01
(68)	ehksaxlaxnt	7	0.09	(72)	bihgihst	3	0.04	(74)	behst	1	0.01
(68)	aem	7	0.09	(72)	frahnt	3	0.04	(74)	laajhihst	1	0.01
(68)	lihtaxl	7	0.09	(72)	wohz	3	0.04	(74)	iyvaxn	1	0.01
(68)	slaytlih	7	0.09	(72)	biy	3	0.04	(74)	showd	1	0.01
(68)	ow	7	0.09	(72)	stihl	3	0.04	(74)	dhaxz	1	0.01
(68)	dheaz	7	0.09	(72)	seym	3	0.04	(74)	mowmaxnt	1	0.01
(68)	hhay	7	0.09	(72)	maonihng	3	0.04	(74)	hh	1	0.01
(68)	blohk	7	0.09	(72)	taym	3	0.04	(74)	stohp	1	0.01
(68)	sayz	7	0.09	(72)	aant	3	0.04	(74)	erm	1	0.01
(68)	iych	7	0.09	(72)	skweaz	3	0.04	(74)	rihng	1	0.01
(69)	luhks	6	0.07	(72)	serkaxlz	3	0.04	(74)	hhaed	1	0.01
(69)	yua	6	0.07	(72)	kwayt	3	0.04	(74)	krohsihz	1	0.01
(69)	dawnwaxdz	6	0.07	(73)	gehtihng	2	0.02	(74)	rehkaxgnayz	1	0.01
(69)	bay	6	0.07	(73)	wuhd	2	0.02	(74)	kohpihihng	1	0.01
(69)	rialih	6	0.07	(73)	saynz	2	0.02	(74)	tehl	1	0.01

Table B.9 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(74)	trayaenggaxl	1	0.01	(74)	aolweyz	1	0.01	(74)	ehnih	1	0.01
(74)	neym	1	0.01	(74)	pihkchaxz	1	0.01	(74)	hhowld	1	0.01
(74)	hhaaf	1	0.01	(74)	greyt	1	0.01	(74)	gihv	1	0.01
(74)	kahlaxd	1	0.01	(74)	loht	1	0.01	(74)	dihsayd	1	0.01
(74)	sehkaxnd	1	0.01	(74)	brayt	1	0.01	(74)	kaol	1	0.01
(74)	klehvax	1	0.01	(74)	taynih	1	0.01	(74)	gehsihng	1	0.01
(74)	tiyzihng	1	0.01	(74)	siym	1	0.01	(74)	ihksaytihd	1	0.01
(74)	hhiyz	1	0.01	(74)	prihtih	1	0.01	(74)	wohts	1	0.01
(74)	ahp	1	0.01	(74)	rihmehmbax	1	0.01	(74)	krsaxnt	1	0.01
(74)	axrawnd	1	0.01	(74)	aekchualih	1	0.01	(74)	sayn	1	0.01
(74)	dahz	1	0.01	(74)	hhiaz	1	0.01	(74)	miynz	1	0.01
(74)	hhowl	1	0.01	(74)	may	1	0.01	(74)	taokaxtihv	1	0.01
(74)	cheynjhihng	1	0.01	(74)	nehvax	1	0.01	(74)	mihsihng	1	0.01
(74)	dhey	1	0.01	(74)	paetaxn	1	0.01	(74)	blohks	1	0.01
(74)	gohn	1	0.01	(74)	thihngz	1	0.01	(74)	chaetih	1	0.01
(74)	biht	1	0.01	(74)	miyn	1	0.01	(74)	sehntax	1	0.01
(74)	faynaxlih	1	0.01	(74)	hhaepih	1	0.01	(74)	fahsih	1	0.01
(74)	krihstiyn	1	0.01	(74)	kyuwbz	1	0.01	(74)	wer	1	0.01
(74)	taokihng	1	0.01	(74)	dihskrayb	1	0.01				

Table B.10: ***Complete word-frequencies of all words in the experiment of Saunders et al. (2012)****. Listed are the rank, word count (*cnt*) and the percentage relative to the total number of words in the experiment across all participants and sessions.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	bluw	157	6.91	(17)	sahn	28	1.23	(29)	yehs	13	0.57
(2)	rehd	126	5.54	(18)	wahn	27	1.19	(29)	smaolax	13	0.57
(3)	serkaxl	117	5.15	(19)	kahlax	24	1.06	(29)	krehsaxnt	13	0.57
(4)	hhaat	108	4.75	(20)	guhd	22	0.97	(30)	miydiam	12	0.53
(5)	griyn	99	4.36	(20)	bihgax	22	0.97	(31)	kaxrehkt	11	0.48
(6)	aerow	81	3.56	(21)	ax	21	0.92	(31)	layk	11	0.48
(7)	krohs	79	3.48	(22)	dhaets	20	0.88	(31)	yea	11	0.48
(7)	sayd	79	3.48	(23)	diychiy	19	0.84	(32)	brihliant	10	0.44
(8)	bohks	64	2.82	(24)	sheyps	18	0.79	(32)	now	10	0.44
(9)	sheyp	55	2.42	(24)	iht	18	0.79	(32)	ihts	10	0.44
(10)	aend	48	2.11	(24)	baekgrawnd	18	0.79	(32)	naw	10	0.44
(10)	muwn	48	2.11	(24)	yuw	18	0.79	(32)	dahn	10	0.44
(11)	skwea	47	2.07	(25)	smayl	17	0.75	(33)	mihdaxl	9	0.4
(12)	dhihs	46	2.02	(25)	dhaet	17	0.75	(34)	dhax	8	0.35
(13)	staa	42	1.85	(25)	baodax	17	0.75	(34)	taxdey	8	0.35
(14)	uhkey	40	1.76	(25)	wayt	17	0.75	(34)	aolsow	8	0.35
(14)	ihz	40	1.76	(26)	goht	16	0.7	(34)	luhk	8	0.35
(15)	smaol	35	1.54	(26)	laajh	16	0.7	(34)	ohn	8	0.35
(15)	rayt	35	1.54	(27)	axbawt	15	0.66	(34)	hhaxlow	8	0.35
(16)	hhia	29	1.28	(28)	woht	14	0.62	(35)	saydz	7	0.31

Table B.10 – *Continued from previous page*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(35)	bohksihz	7	0.31	(39)	smaolihst	3	0.13	(41)	behst	1	0.04
(35)	axgehn	7	0.31	(39)	sahm	3	0.13	(41)	rihmehmbax	1	0.04
(35)	wehl	7	0.31	(39)	nays	3	0.13	(41)	show	1	0.04
(35)	er	7	0.31	(39)	slaytlih	3	0.13	(41)	aekchualih	1	0.04
(36)	waots	6	0.26	(39)	aet	3	0.13	(41)	hhaw	1	0.04
(36)	dihfraxnt	6	0.26	(39)	ow	3	0.13	(41)	ihl	1	0.04
(36)	downt	6	0.26	(39)	seym	3	0.13	(41)	faynd	1	0.04
(36)	ahpwaxdz	6	0.26	(40)	maonihng	2	0.09	(41)	aant	1	0.04
(36)	ihznt	6	0.26	(40)	wahnz	2	0.09	(41)	showd	1	0.04
(36)	gow	6	0.26	(40)	poyntihng	2	0.09	(41)	showz	1	0.04
(36)	siy	6	0.26	(40)	ihgzaampaxl	2	0.09	(41)	sihtihng	1	0.04
(36)	bihg	6	0.26	(40)	dawn	2	0.09	(41)	kaen	1	0.04
(36)	dhea	6	0.26	(40)	wey	2	0.09	(41)	hh	1	0.04
(37)	ehksaxlaxnt	5	0.22	(40)	yp	2	0.09	(41)	ohv	1	0.04
(37)	showihng	5	0.22	(40)	skweaz	2	0.09	(41)	aol	1	0.04
(37)	tray	5	0.22	(40)	bihgihst	2	0.09	(41)	sahmthihng	1	0.04
(37)	uwm	5	0.22	(40)	sihks	2	0.09	(41)	rihng	1	0.04
(37)	kahlaxz	5	0.22	(40)	bihfao	2	0.09	(41)	krohsihz	1	0.04
(38)	kyuwb	4	0.18	(40)	kwayt	2	0.09	(41)	dihskrayb	1	0.04
(38)	luhkihng	4	0.18	(40)	thihngk	2	0.09	(41)	dhaen	1	0.04
(38)	tuw	4	0.18	(40)	sahmtaymz	2	0.09	(41)	rehkaxgnayz	1	0.04
(38)	rawnd	4	0.18	(40)	dhehn	2	0.09	(41)	hhaaf	1	0.04
(38)	ihn	4	0.18	(40)	ferst	2	0.09	(41)	frahnt	1	0.04
(38)	wihdh	4	0.18	(40)	wohnt	2	0.09	(41)	kahlaxd	1	0.04
(38)	blohk	4	0.18	(40)	ihnsayd	2	0.09	(41)	sehkaxnd	1	0.04
(38)	lihtaxl	4	0.18	(40)	ihm	2	0.09	(41)	gehsihng	1	0.04
(38)	sow	4	0.18	(40)	dawnwaxdz	2	0.09	(41)	tiyzihng	1	0.04
(38)	vehrih	4	0.18	(41)	aolrehdih	1	0.04	(41)	wohz	1	0.04
(39)	hhaev	3	0.13	(41)	ao	1	0.04	(41)	ihksaytihd	1	0.04
(39)	sfia	3	0.13	(41)	tern	1	0.04	(41)	ahp	1	0.04
(39)	dihdnt	3	0.13	(41)	faynaxlih	1	0.04	(41)	axrawnd	1	0.04
(39)	laast	3	0.13	(41)	ay	1	0.04	(41)	wohts	1	0.04
(39)	noht	3	0.13	(41)	wohchihng	1	0.04	(41)	krsaxnt	1	0.04
(39)	aa	3	0.13	(41)	krihstiyn	1	0.04	(41)	wihch	1	0.04
(39)	baht	3	0.13	(41)	aolweyz	1	0.04	(41)	taokaxtihv	1	0.04
(39)	luhks	3	0.13	(41)	lahvlih	1	0.04	(41)	cheynjhihng	1	0.04
(39)	feyvaxriht	3	0.13	(41)	geht	1	0.04	(41)	sayz	1	0.04
(39)	rialih	3	0.13	(41)	owkey	1	0.04	(41)	feysihng	1	0.04
(39)	hhaez	3	0.13	(41)	pihkchaxz	1	0.04	(41)	nehkst	1	0.04
(39)	serkaxlz	3	0.13	(41)	bay	1	0.04	(41)	saxrawndihd	1	0.04
(39)	lehts	3	0.13	(41)	greyt	1	0.04	(41)	sehntax	1	0.04
(39)	paetaxnz	3	0.13	(41)	kaent	1	0.04	(41)	dhiyz	1	0.04
(39)	axnahdhax	3	0.13	(41)	brayt	1	0.04	(41)	laajhax	1	0.04
(39)	sayzd	3	0.13	(41)	skay	1	0.04				

B.4 Accumulated word frequencies for the first three sessions

Table B.11: ***Word-frequencies of all words of the first three sessions in rejection experiment****. Listed are the ten most frequent words within said experiment produced during the first three sessions across all participants. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	you	772	7.16	*(34)*	yes	84	0.78	*(81)*	didn't (2)	10	0.09
(2)	the	590	5.47	*(36)*	Deechee	76	0.7	*(84)*	didn't	7	0.06
(3)	like	374	3.47	*(39)*	not	71	0.66	*(85)*	isn't	6	0.06
(4)	a	346	3.21	*(44)*	don't	56	0.52	*(85)*	moons	6	0.06
(5)	one	287	2.66	*(45)*	box	55	0.51	*(85)*	rectangles	6	0.06
(6)	this	281	2.6	*(45)*	triangles	55	0.51	*(87)*	pyramid	4	0.04
(7)	no	235	2.18	*(52)*	Deechee (2)	45	0.42	*(88)*	smile	3	0.03
(9)	square	210	1.95	*(55)*	crescent	40	0.37	*(88)*	pyramids	3	0.03
(10)	do	206	1.91	*(60)*	shape	32	0.3	*(88)*	haven't	3	0.03
(12)	to	183	1.7	*(62)*	nice	29	0.27	*(88)*	smiling	3	0.03
(13)	heart	177	1.64	*(66)*	favourite	25	0.23	*(88)*	won't	3	0.03
(14)	moon	174	1.61	*(68)*	happy	23	0.21	*(89)*	can't	2	0.02
(16)	triangle	152	1.41	*(73)*	hearts	18	0.17	*(89)*	doesn't (2)	2	0.02
(18)	circle	139	1.29	*(74)*	target	17	0.16	*(90)*	wouldn't	1	0.01
(20)	don't	128	1.19	*(76)*	sad	15	0.14	*(90)*	doesn't	1	0.01
(23)	circles	116	1.08	*(77)*	arteen	14	0.13	*(90)*	weren't	1	0.01
(25)	squares	111	1.03	*(80)*	know	11	0.1				

Table B.12: ***Word-frequencies of all words of the first three sessions in prohibition experiment.*** *Listed are the ten most frequent words within said experiment produced during the first three sessions across all participants. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	you	918	5.98	*(28)*	don't	129	0.84	*(92)*	smile	7	0.05
(2)	the	788	5.13	*(29)*	play	119	0.77	*(93)*	doesn't	6	0.04
(3)	a	659	4.29	*(33)*	circles	102	0.66	*(94)*	mustn't	5	0.03
(4)	this	572	3.73	*(34)*	crescent	100	0.65	*(94)*	sad	5	0.03
(5)	one	459	2.99	*(39)*	yes	91	0.59	*(94)*	smiling	5	0.03
(6)	no	352	2.29	*(39)*	can't (2)	91	0.59	*(94)*	haven't	5	0.03
(7)	is	347	2.26	*(46)*	box	74	0.48	*(96)*	won't	3	0.02
(8)	like	298	1.94	*(53)*	Deechee	59	0.38	*(96)*	hasn't	3	0.02
(9)	it's	282	1.84	*(53)*	Deechee (2)	59	0.38	*(97)*	cannot	2	0.01
(10)	to	269	1.75	*(62)*	squares	44	0.29	*(98)*	moons	1	0.01
(11)	heart	236	1.54	*(66)*	nice	37	0.24	*(98)*	nono	1	0.01
(13)	square	224	1.46	*(66)*	triangles	37	0.24	*(98)*	target	1	0.01
(15)	triangle	214	1.39	*(69)*	hearts	34	0.22	*(98)*	crescents	1	0.01
(16)	moon	211	1.37	*(78)*	know	22	0.14	*(98)*	wouldn't	1	0.01
(17)	shape	201	1.31	*(82)*	happy	17	0.11	*(98)*	neither	1	0.01
(20)	circle	180	1.17	*(85)*	favourite	14	0.09	*(98)*	shouldn't	1	0.01
(21)	do	175	1.14	*(88)*	isn't	11	0.07				
(22)	not	170	1.11	*(90)*	didn't	9	0.06				

Table B.13: ***Word-frequencies of salient words of the first three sessions in rejection experiment.*** *Listed are the ten most frequent salient words within said experiment produced during the first three sessions across all participants. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot.*

rank	word	cnt	%	*rank*	word	cnt	%	*rank*	word	cnt	%
(1)	square	157	4.99	*(23)*	don't	32	1.02	*(40)*	happy	11	0.35
(2)	no	146	4.64	*(26)*	Deechee (2)	27	0.86	*(41)*	a	10	0.32
(3)	triangle	127	4.04	*(27)*	box	26	0.83	*(45)*	do	6	0.19
(4)	heart	122	3.88	*(28)*	crescent	25	0.8	*(46)*	rectangles	5	0.16
(5)	circle	111	3.53	*(31)*	are	20	0.64	*(47)*	isn't	4	0.13
(6)	moon	110	3.5	*(34)*	shape	17	0.54	*(47)*	pyramid	4	0.13
(7)	like	108	3.44	*(35)*	favourite	16	0.51	*(48)*	smile	3	0.1
(8)	circles	88	2.8	*(36)*	target	15	0.48	*(48)*	pyramids	3	0.1
(9)	one	84	2.67	*(36)*	nice	15	0.48	*(48)*	moons	3	0.1
(10)	squares	75	2.39	*(38)*	not	13	0.41	*(49)*	smiling	2	0.06
(11)	it	71	2.26	*(39)*	hearts	12	0.38	*(49)*	won't	2	0.06
(13)	yes	56	1.78	*(39)*	arteen	12	0.38	*(50)*	didn't (2)	1	0.03
(16)	Deechee	46	1.46	*(39)*	sad	12	0.38	*(50)*	can't	1	0.03
(17)	this	42	1.34	*(40)*	the	11	0.35	*(50)*	didn't	1	0.03
(22)	triangles	33	1.05	*(40)*	don't (2)	11	0.35				
(23)	you	32	1.02	*(40)*	to	11	0.35				

Table B.14: ***Word-frequencies of salient words of the first three sessions in prohibition experiment.*** *Listed are the ten most frequent salient words within said experiment produced during the first three sessions across all participants. Given are the rank, the word count (*cnt*) and the percentage relative to the total number of words in the experiment. Apart from the highest-ranking words the same statistics are given for object labels, negation words, and words linked to the motivational state of the robot.*

rank	word	cnt	%	*(rank)*	word	cnt	%	*(rank)*	word	cnt	%
(1)	you	918	5.98	*(28)*	don't	129	0.84	*(92)*	smile	7	0.05
(2)	the	788	5.13	*(29)*	play	119	0.77	*(93)*	doesn't	6	0.04
(3)	a	659	4.29	*(33)*	circles	102	0.66	*(94)*	mustn't	5	0.03
(4)	this	572	3.73	*(34)*	crescent	100	0.65	*(94)*	sad	5	0.03
(5)	one	459	2.99	*(39)*	yes	91	0.59	*(94)*	smiling	5	0.03
(6)	no	352	2.29	*(39)*	can't (2)	91	0.59	*(94)*	haven't	5	0.03
(7)	is	347	2.26	*(46)*	box	74	0.48	*(96)*	won't	3	0.02
(8)	like	298	1.94	*(53)*	Deechee	59	0.38	*(96)*	hasn't	3	0.02
(9)	it's	282	1.84	*(53)*	Deechee (2)	59	0.38	*(97)*	cannot	2	0.01
(10)	to	269	1.75	*(62)*	squares	44	0.29	*(98)*	moons	1	0.01
(11)	heart	236	1.54	*(66)*	nice	37	0.24	*(98)*	nono	1	0.01
(13)	square	224	1.46	*(66)*	triangles	37	0.24	*(98)*	target	1	0.01
(15)	triangle	214	1.39	*(69)*	hearts	34	0.22	*(98)*	crescents	1	0.01
(16)	moon	211	1.37	*(78)*	know	22	0.14	*(98)*	wouldn't	1	0.01
(17)	shape	201	1.31	*(82)*	happy	17	0.11	*(98)*	neither	1	0.01
(20)	circle	180	1.17	*(85)*	favourite	14	0.09	*(98)*	shouldn't	1	0.01
(21)	do	175	1.14	*(88)*	isn't	11	0.07				
(22)	not	170	1.11	*(90)*	didn't	9	0.06				

B.5 Listings of In-group changes between sessions of utterance level measures

B.5.1 Rejection Experiment

Table B.15: *Percental changes of **MLU** between sessions of rejection scenario. Statistically significant changes ($p < .05$) are in bold*

(a) *All Utterances, T: ∗=1.98, ∗∗=2.37, †=2.38, ‡=2.9*

	s1→s2	s2→s3	s3→s4	s4→s5
P01	-4.55	14.29	4.17	4
P04	-10.64	0	0	-4.76
P05	3.57	3.45	-6.67	-7.14
P06	-2.63	5.41	-5.13	8.11
P07	12.5	**−11.11***	-9.38	6.9
P08	15.63	-8.11	2.94	-8.57
P09	-5.26	2.78	-5.41	-14.29
P10	18.75	5.26	25	-28
P11	**−20****	**−21.88**†	8	7.41
P12	**−26.32**‡	0	-17.86	8.7

(b) *Negative Utterances Only, T: [1]:2.69, [2]:2.43, [3]:2.7, [4] : 4.4, [5]:2.37, [6]:2.27,[7]:2.15*

	s1→s2	s2→s3	s3→s4	s4→s5
P01	35	25.93	**−47.06**[1]	**105.56**[2]
P04	11.36	8.16	-13.21	2.17
P05	**66.67**[3]	**−60**[4]	166.67[5]	-9.38
P06	-5.41	31.43	-15.22	12.82
P07	28.13	-26.83	6.67	-15.63
P08	28.57	**−48.89**[6]	39.13	21.88
P09	37.78	0	-17.74	**−33.33**[7]
P10	inf	20	56.67	-57.45
P11	-25	4.76	-22.73	8.82
P12	-26.32	42.86	-35	26.92

Table B.16: *Percental changes of **distinct words** between sessions of rejection scenario; a statistical evaluation is not possible for the depicted values as each is based on singular values for each session and participant*

(a) *All Utterances*

	s1→s2	s2→s3	s3→s4	s4→s5
P01	23.81	11.54	3.45	-23.33
P04	-23.76	38.96	-2.8	0
P05	-26.04	2.82	-8.22	20.9
P06	-14	12.79	-2.06	-6.32
P07	-23.57	-6.54	8	-2.78
P08	-38.62	7.87	-8.33	44.32
P09	3.76	12.32	-1.94	-23.68
P10	350	0	-14.81	-13.04
P11	-12.62	-37.78	3.57	27.59
P12	-33.33	-71.05	22.73	33.33

(b) *Negative Utterances Only*

Due to the small number of distinct negative words, percentual values were not calculated for negative words only.

See table 5.10 for absolute values.

Table B.17: *Percental changes of **utterances per minute (u/min)** between sessions of rejection scenario.* undef *means that there were 0 utterances in session X, for X→Y.*

(a) *All Utterances*

	s1→s2	s2→s3	s3→s4	s4→s5
P01	26.36	-7.19	16.28	-20
P04	6.28	-11.81	-6.25	19.05
P05	-16.6	13.43	2.04	20.4
P06	-8.86	-0.63	7.26	-9.71
P07	2.26	-3.04	2.28	0
P08	-20.06	0	-19.31	13.4
P09	-4.4	16.74	2.51	-8.04
P10	114.29	-14.67	6.25	8.82
P11	-6.99	-9.86	-27.08	-0.71
P12	-14.67	-52.23	76	-18.18

(b) *Negative Utterances Only*

	s1→s2	s2→s3	s3→s4	s4→s5
P01	200	66.67	50	-60
P04	34.88	-20.69	-2.17	68.89
P05	-17.02	-38.46	41.67	5.88
P06	-26.47	28	-3.13	-25.81
P07	52.27	1.49	-30.88	61.7
P08	-9.3	-23.08	33.33	-50
P09	-2.78	14.29	-30	7.14
P10	undef	0	50	0
P11	38.1	-50	37.93	-72.5
P12	3.85	-40.74	75	-7.14

B.5.2 Prohibition Experiment

Table B.18: *Percental changes of **MLU** between sessions of prohibition scenario. Statistically significant changes ($p < .05$) are typed bold*

(a) *All Utterances, T:* [1]*=2.09,* [2]*=2.53,* [3]*=2.23,* [4]*=2.49,* [5]*=2.2*

	s1→s2	s2→s3	s3→s4	s4→s5
P13	-7.5	**−13.51**[1]	3.12	-9.09
P14	10	9.09	-8.33	3.03
P15	2.7	5.26	-2.5	-10.26
P16	13.89	**−14.63**[2]	5.71	5.41
P17	-6.67	-7.14	19.23	-3.23
P18	0	-2.7	5.56	-5.26
P19	0	0	-5.56	-2.94
P20	**18.92**[3]	4.55	**−19.57**[4]	0
P21	-3.57	-14.81	**21.74**[5]	-7.14
P22	-9.09	-3.33	0	0

(b) *Negative Utterances Only, T:* [1]*=2.95,* [2]*=2.11,* [3]*=2.15*

	s1→s2	s2→s3	s3→s4	s4→s5
P13	-8.33	0.00	-13.64	39.47
P14	undef	25	-66.67	30
P15	15.56	-3.85	0	**−44**[1]
P16	-14.29	4.17	10	-3.64
P17	-22.22	-9.52	0	-31.58
P18	6.82	4.26	4.08	-9.8
P19	15.79	6.82	0	-8.51
P20	**46.15**[2]	-7.02	-18.87	4.65
P21	-13.89	-22.58	**66.67**[3]	-37.5
P22	-35.42	-12.9	44.44	-28.21

Table B.19: *Percental changes of **distinct words** between sessions of prohibition scenario*

(a) *All Utterances*

	s1→s2	s2→s3	s3→s4	s4→s5
P13	-9.09	-30	-1.43	-7.25
P14	-3.95	-4.11	-21.43	-5.45
P15	-5.97	20.63	-13.16	-0.76
P16	-4.1	9.63	-4.88	1.54
P17	-15.87	-1.89	-23.08	5.00
P18	10.59	-5.32	8.99	-9.28
P19	-34.27	11.11	-10	-6.84
P20	-3.51	-20.91	6.90	12.90
P21	1.35	9.33	-28.05	16.95
P22	-22.22	29.87	27	-16.54

(b) *Negative Utterances Only*

Due to the small number of distinct negative words, percentual values were not calculated for negative words only.

See table 5.11 for absolute values.

Table B.20: *Percental changes of* ***utterances per minute (u/min)*** *between sessions of prohibition scenario*

(a) *All Utterances*

	s1→s2	s2→s3	s3→s4	s4→s5
P13	23.68	-18.54	11.94	4.33
P14	5.17	-6.56	-2.34	-5.39
P15	16.98	-1.34	-4.09	1.42
P16	14	1.75	-10.34	-3.3
P17	16.15	-6.62	12.06	32.28
P18	3.25	-0.52	-0.53	4.77
P19	-15.13	12.94	-1.55	7.86
P20	-2.54	1.12	-3.31	18.25
P21	26.13	9.96	-14.13	5.06
P22	-9.44	45.97	25.65	-17.05

(b) *Negative Utterances Only*

	s1→s2	s2→s3	s3→s4	s4→s5
P13	76.32	-79.1	-28.57	20
P14	undef	-12.5	-85.71	150
P15	-26	24.32	-48.91	23.4
P16	40	-19.05	-7.84	8.51
P17	10.71	61.29	-70	-6.67
P18	-50	25	-40	-18.52
P19	23.68	0	-74.47	41.67
P20	26.09	1.72	-32.2	0
P21	36.36	4.44	-65.96	93.75
P22	262.5	17.24	-61.76	-15.38

B.5.3 Saunders' Experiment

Table B.21: *Percental changes of **MLU** between sessions of Saunders' experiment*

(a) *All Utterances*

	s1→s2	s2→s3	s3→s4	s4→s5
M02	0	3.23	-18.75	15.38
F05	26.47	-11.63	13.16	-9.3
M03	10.34	-15.63	-3.7	7.69
F01	7.89	-12.2	5.56	15.79
F02	-8.16	0	-2.22	6.82
M01	2.94	11.43	-15.38	-12.12
F03	n/a	-11.43	6.45	-6.06
F06	-8.11	11.76	2.63	-10.26
F04	-23.08	6.67	0	-18.75

(b) *Negative Utterances Only*

Due to the small number of or even absent negative words, percentual values were not calculated for negative words only.

See table B.4 for absolute values.

Table B.22: *Percental changes in number of **distinct words** between sessions of Saunders' experiment*

(a) *All Utterances*

	s1→s2	s2→s3	s3→s4	s4→s5
M02	-17.24	-12.5	-14.29	33.33
F05	10	-25.97	-22.81	9.09
M03	8.82	5.41	-25.64	27.59
F01	-11.9	-27.03	7.41	0
F02	-12.16	-12.31	-24.56	4.65
M01	-24.14	-18.18	-30.56	-8
F03		19.64	-1.49	-13.64
F06	-19.42	-4.82	8.86	-3.49
F04	-51.76	48.78	-14.75	-11.54

(b) *Negative Utterances Only*

Due to the small number of or even absent negative words, percentual values were not calculated for negative words only.

See table B.4 for absolute values.

Table B.23: *Percental changes in* ***utterances per minute (u/min)*** *between sessions of Saunders' experiment*

(a) *All Utterances*

	s1→s2	s2→s3	s3→s4	s4→s5
M02	12.43	-21.61	7.05	-7.19
F05	-14.67	19.46	-23.11	29.06
M03	56.55	-17.62	-14.44	5
F01	-6.32	0.56	-18.99	2.07
F02	-8.04	-3.42	-24.8	14.66
M01	24.32	-29.35	-7.69	3.89
F03	n/a	11.07	0.96	11.15
F06	-0.6	-1.5	-0.3	-4.28
F04	-19.73	6.98	-11.8	-1.41

(b) *Negative Utterances Only*

Due to the small number of or even absent negative words, percentual values were not calculated for negative words only.

See table B.4 for absolute values.

B.6 Negative Words vs. Negation Types

The following tables display the numerical basis based on which the statistical analyses presented in section 5.3.9, subsection 'Pragmatic Analysis of Participants' (pp. 280), were performed. Table B.24 presents the absolute counts of the four most frequent negation words - *no*, *don't*, *not*, and *can't* - grouped by negation type, i.e. the frequencies with which the respective negation words were produced within utterances that were classified as instances of the stated negation types.

Tables B.25 and B.26 show the relative frequencies (percentages) of the respective negation word relative to the total number of the displayed four negation words produced within utterances categorised as instances of the displayed negation type. All displayed percentages were rounded to two decimal places for purposes of display, yet the actual computation was performed on the same numbers with four decimal places. Example: Participant *P04* produced 29 of the four most frequent negative words within utterances which were categorized as being instances of *negative intent interpretations*. 13, that is 44.82%, of these words were *no*'s. If we subsequently assert that 44.82% of the negation words associated with *negative intent interpretations* are *no*'s we commit a small error. This error is due to us only considering the four most frequent negation words and disregarding the less frequent types. On the other hand we know that our most frequent 4 negation words cover for 97.8% of all negation words that were produced by all participants in utterances classified as being of this type (cf. table 5.46). Therefore the error in our calculation is relatively small. Similarly high coverage rates apply for *negative motivational questions*, *prohibitions*, *disallowances*, and *truth-functional denials*, with the lowest of these rates being 97.3% in the case of *negative motivational questions*. Subsequently the error caused by ignoring negation words other than the most frequent four ones is marginal. Yet

we cannot interpret these percentages to mean "44.82% of utterances that are instances of *negative motivational questions* contain a *no*" because there is no 1 : 1 relation between negation words, negative utterances, and negation types: some of the utterances based upon which these percentages were calculated contain more than one negation word, for example "No, you don't like the star". Some of these utterances even contain the same negation word several times such as "no no no no".

Tables B.27 and B.28 show the salience rates in percent of the corresponding negation words, grouped by type and participant. These rates were calculated relative to the absolute counts given in table B.24 and under exclusion of the same counts as indicated in the tables B.25 and B.26 (entry: *n/a*). Naturally there are neither salience rates stated (*n/a*-entry) for those cases in which a particular participant never produced the respective word in conjunction with the respective type. Example: The entry for participant *P08*, the word *not*, and the negation type *negative motivational questions* is 0.2. This means that 20% of the 5 productions (= 1 production) of *not* by *P08* within utterances that were classified as *negative motivational questions* were such that *not* was extracted as the salient word of the utterance.

Table B.24: ***Absolute frequencies of most frequent negation words grouped negation type.*** *Displayed is the distribution of the most frequent negation words grouped by the most frequent negation types within instances of which they were produced. For relative frequencies (percentages) see table B.25. Abbreviations:* ***nWord****: negation word,* ***I****: negative intent interpretations,* ***Q****: negative motivational questions,* ***P****: prohibition,* ***D****: disallowance,* ***T****: truth-functional denial*

nWord	type	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
	I	0	13	18	16	17	9	5	1	11	7	3	0	28	12	7	9	7	11	1	0
	Q	0	22	11	3	52	26	2	7	1	3	2	0	29	1	7	3	2	13	7	0
no	P	0	0	0	0	0	0	0	0	0	0	10	12	24	11	29	1	3	15	21	3
	D	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	1	2	33	0
	T	14	24	2	1	2	0	7	0	46	41	11	2	1	0	6	1	14	0	6	33
	I	0	1	7	7	4	0	5	0	1	1	3	0	1	8	2	11	5	2	0	1
	Q	0	3	6	0	3	5	4	0	0	0	0	0	8	3	0	5	1	7	2	0
not	P	0	0	0	0	0	0	0	0	0	0	0	1	17	1	1	3	1	8	3	5
	D	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	2	0	13	5	0
	T	4	24	0	0	2	0	3	0	12	7	13	1	2	0	2	1	14	4	1	3
	I	2	15	14	27	31	9	16	0	10	3	11	0	14	14	4	11	7	10	3	1
	Q	0	7	10	7	17	15	15	2	6	1	7	0	21	10	0	6	9	19	11	0
don't	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
	D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
can't	P	0	0	0	0	0	0	0	0	0	0	12	6	10	11	3	9	11	5	0	1
	D	0	0	0	0	0	0	0	0	0	0	0	0	8	2	0	1	3	1	0	1
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table B.25: ***Relative frequencies of most frequent negation words in relation to negation types.*** *Displayed is the distribution of the most frequent negation words which were produced as part of instances of the most frequent negation types in percentages relative to the total number of negation words of the respective types. Percentages overwritten with* n/a *were excluded due to too small of an underlying numerical basis (#* ***instances*** *<* **5***). Abbreviations:* ***nWord****: negation word,* ***I****: negative intent interpretations,* ***Q****: negative motivational questions,* ***P****: prohibition,* ***D****: disallowance,* ***T****: truth-functional denial.*

nWord	type	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
	I	n/a	0.44	0.46	0.32	0.32	0.5	0.19	n/a	0.5	0.63	0.17	n/a	0.65	0.35	0.53	0.29	0.36	0.47	n/a	n/a
	Q	n/a	0.68	0.40	0.3	0.72	0.56	0.09	0.77	0.14	n/a	0.22	n/a	0.5	0.07	1	0.21	0.17	0.33	0.35	n/a
no	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.45	0.63	0.47	0.48	0.88	0.08	0.2	0.54	0.81	0.33
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.07	n/a	n/a	n/a	n/a	0.13	0.83	n/a
	T	0.78	0.5	n/a	n/a	n/a	n/a	0.7	n/a	0.79	0.85	0.46	n/a	n/a	n/a	0.75	n/a	0.5	n/a	0.86	0.92
	I	n/a	0.03	0.18	0.14	0.08	0	0.19	n/a	0.05	0.09	0.18	n/a	0.02	0.24	0.15	0.35	0.26	0.09	n/a	n/a
	Q	n/a	0.09	0.22	0	0.04	0.11	0.19	0	0	n/a	0	n/a	0.14	0.21	0	0.36	0.08	0.18	0.1	n/a
not	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0.05	0.33	0.04	0.03	0.23	0.07	0.29	0.12	0.56
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.4	n/a	n/a	n/a	n/a	0.81	0.13	n/a
	T	0.22	0.5	n/a	n/a	n/a	n/a	0.3	n/a	0.21	0.15	0.54	n/a	n/a	n/a	0.25	n/a	0.5	n/a	0.14	0.08
	I	n/a	0.52	0.36	0.54	0.60	0.5	0.62	n/a	0.45	0.27	0.65	n/a	0.33	0.41	0.31	0.35	0.37	0.43	n/a	n/a
	Q	n/a	0.22	0.37	0.7	0.24	0.33	0.71	0.22	0.86	n/a	0.78	n/a	0.36	0.71	0	0.43	0.75	0.49	0.55	n/a
don't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0	0	0	0.08	0
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	n/a	n/a	n/a	n/a	0	0.05	n/a
	T	0	0	n/a	n/a	n/a	n/a	0	n/a	0	0	0	n/a	n/a	n/a	0	n/a	0	n/a	0	0
	I	n/a	0	0	0	0	0	0	n/a	0	0	0	n/a	0	0	0	0	0	0	n/a	n/a
	Q	n/a	0	0	0	0	0	0	0	0	n/a	0	n/a	0	0	0	0	0	0	0	n/a
can't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.55	0.32	0.20	0.48	0.09	0.69	0.73	0.18	0	0.11
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.53	n/a	n/a	n/a	n/a	0.06	0	n/a
	T	0	0	n/a	n/a	n/a	n/a	0	n/a	0	0	0	n/a	n/a	n/a	0	n/a	0	n/a	0	0

Table B.26: ***Relative frequencies of most frequent negation words in relation to negation types.*** *Displayed is the distribution of the most frequent negation words which were produced as part of instances of the most frequent negation types in percentages relative to the total number of negation words of the respective types. Percentages overwritten with* n/a *were excluded due to too small of an underlying numerical basis (#* ***instances*** < 20*). Abbreviations:* ***nWord****: negation word,* ***I****: negative intent interpretations,* ***Q****: negative motivational questions,* ***P****: prohibition,* ***D****: disallowance,* ***T****: truth-functional denial.*

nWord	type	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
	I	n/a	0.45	0.46	0.32	0.33	n/a	0.19	n/a	n/a	n/a	n/a	n/a	0.65	0.35	0.54	0.29	n/a	0.48	n/a	n/a
	Q	n/a	0.69	0.41	n/a	0.72	0.57	0.10	n/a	n/a	n/a	n/a	n/a	0.5	n/a	n/a	n/a	n/a	0.33	0.35	n/a
no	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.45	n/a	0.47	n/a	0.88	n/a	n/a	0.54	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.83	n/a
	T	n/a	0.5	n/a	n/a	n/a	n/a	n/a	n/a	0.80	0.85	0.46	n/a	n/a	n/a	n/a	n/a	0.5	n/a	n/a	0.92
	I	n/a	0.03	0.18	0.14	0.08	n/a	0.19	n/a	n/a	n/a	n/a	n/a	0.02	0.24	0.15	0.35	n/a	0.09	n/a	n/a
	Q	n/a	0.09	0.22	n/a	0.04	0.11	0.19	n/a	n/a	n/a	n/a	n/a	0.14	n/a	n/a	n/a	n/a	0.18	0.1	n/a
not	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	n/a	0.33	n/a	0.03	n/a	n/a	0.29	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.13	n/a
	T	n/a	0.5	n/a	n/a	n/a	n/a	n/a	n/a	0.21	0.15	0.54	n/a	n/a	n/a	n/a	n/a	0.5	n/a	n/a	0.08
	I	n/a	0.52	0.36	0.54	0.60	n/a	0.62	n/a	n/a	n/a	n/a	n/a	0.33	0.41	0.31	0.35	n/a	0.43	n/a	n/a
	Q	n/a	0.22	0.37	n/a	0.24	0.33	0.71	n/a	n/a	n/a	n/a	n/a	0.36	n/a	n/a	n/a	n/a	0.49	0.55	n/a
don't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	n/a	0	n/a	0	n/a	n/a	0	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.05	n/a
	T	n/a	0	n/a	n/a	n/a	n/a	n/a	n/a	0	0	0	n/a	n/a	n/a	n/a	n/a	0	n/a	n/a	0
	I	n/a	0	0	0	0	n/a	0	n/a	n/a	n/a	n/a	n/a	0	0	0	0	n/a	0	n/a	n/a
	Q	n/a	0	0	n/a	0	0	0	n/a	n/a	n/a	n/a	n/a	0	n/a	n/a	n/a	n/a	0	0	n/a
can't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.55	n/a	0.20	n/a	0.09	n/a	n/a	0.18	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	n/a
	T	n/a	0	n/a	n/a	n/a	n/a	n/a	n/a	0	0	0	n/a	n/a	n/a	n/a	n/a	0	n/a	n/a	0

Table B.27: ***Salience rates of most frequent negation words grouped by negation types.*** *Displayed are the salience rates of the most frequent negation words produced within instances of the most frequent negation types in percentages relative to the total number of productions (cf. B.24), grouped per participant and per negation type. Percentages overwritten with* n/a *were excluded due to either too small of an underlying numerical basis of all productions associated with the respective type (#* ***instances*** < 5*) or because no productions of a particular type contained the respective negation word. Abbreviations:* ***nWord****: negation word,* ***I****: negative intent interpretations,* ***Q****: negative motivational questions,* ***P****: prohibition,* ***D****: disallowance,* ***T****: truth-functional denial.*

nWord	type	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
	I	n/a	0.31	0.94	0.63	0.76	0.67	0.6	n/a	0.55	0.43	0.67	n/a	0.5	0.5	0.86	0.78	0.71	0.64	n/a	n/a
	Q	n/a	0.45	1	1	0.87	0.73	1	0.29	1	n/a	1	n/a	0.79	1	0.86	1	1	1	1	n/a
no	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.8	0.83	0.46	0.73	0.76	1	0.67	0.87	0.43	0.33
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1	n/a	n/a	n/a	n/a	0	0.30	n/a
	T	0.57	0.13	n/a	n/a	n/a	n/a	0.71	n/a	0.46	0.15	0.36	n/a	n/a	n/a	0.17	n/a	0.21	n/a	0.33	0.55
	I	n/a	0	0.57	0.43	0.5	n/a	0.6	n/a	0	0	0	n/a	1	0.13	0	0.18	0	0	n/a	n/a
	Q	n/a	0	0.17	n/a	0.33	0.2	0.25	n/a	n/a	n/a	n/a	n/a	0.13	0	n/a	0.2	0	0.14	0.5	n/a
not	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1	0.12	1	0	0	0	0	1	0
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.17	n/a	n/a	n/a	n/a	0	0.6	n/a
	T	0	0.04	n/a	n/a	n/a	n/a	0	n/a	0.08	0	0	n/a	n/a	n/a	0	n/a	0.29	n/a	0	0
	I	n/a	0	0.29	0.41	0.19	0.22	0.25	n/a	0.1	0	0.09	n/a	0.14	0.14	0	0.09	0.29	0.1	n/a	n/a
	Q	n/a	0	0.2	0.57	0.29	0.07	0.2	0	0	n/a	0.14	n/a	0.05	0.2	0	0	0.33	0.05	0.09	n/a
don't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.5	0
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.5	n/a
	T	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Q	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
can't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.25	0.67	0.4	0.55	0	0.44	0.55	0	n/a	0
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.25	n/a	n/a	n/a	n/a	1	n/a	n/a
	T	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table B.28: ***Salience rates of most frequent negation words grouped by negation types.*** *Displayed are the salience rates of the most frequent negation words produced within instances of the most frequent negation types in percentages relative to the total number of productions (cf. B.24), grouped per participant and per negation type. Percentages overwritten with* n/a *were excluded due to either too small of an underlying numerical basis of all productions associated with the respective type (#* ***instances*** $< \mathbf{20}$*) or because no productions of a particular type contained the respective negation word. Abbreviations:* ***nWord****: negation word,* ***I****: negative intent interpretations,* ***Q****: negative motivational questions,* ***P****: prohibition,* ***D****: disallowance,* ***T****: truth-functional denial.*

nWord	type	P01	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
	I	n/a	0.30	0.94	0.63	0.76	n/a	0.6	n/a	n/a	n/a	n/a	n/a	0.5	0.5	0.86	0.78	n/a	0.64	n/a	n/a
	Q	n/a	0.45	1	n/a	0.87	0.73	1	n/a	n/a	n/a	n/a	n/a	0.79	n/a	n/a	n/a	n/a	1	1	n/a
no	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.8	n/a	0.46	n/a	0.76	n/a	n/a	0.87	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.30	n/a
	T	n/a	0.13	n/a	n/a	n/a	n/a	n/a	n/a	0.46	0.15	0.36	n/a	n/a	n/a	n/a	n/a	0.21	n/a	n/a	0.55
	I	n/a	0	0.57	0.43	0.5	n/a	0.6	n/a	n/a	n/a	n/a	n/a	1	0.13	0	0.18	n/a	0	n/a	n/a
	Q	n/a	0	0.17	n/a	0.33	0.2	0.25	n/a	n/a	n/a	n/a	n/a	0.13	n/a	n/a	n/a	n/a	0.14	0.5	n/a
not	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.12	n/a	0	n/a	n/a	0	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.6	n/a
	T	n/a	0.04	n/a	n/a	n/a	n/a	n/a	n/a	0.08	0	0	n/a	n/a	n/a	n/a	n/a	0.29	n/a	n/a	0
	I	n/a	0	0.29	0.41	0.19	n/a	0.25	n/a	n/a	n/a	n/a	n/a	0.14	0.14	0	0.09	n/a	0.1	n/a	n/a
	Q	n/a	0	0.2	n/a	0.29	0.07	0.2	n/a	n/a	n/a	n/a	n/a	0.05	n/a	n/a	n/a	n/a	0.05	0.09	n/a
don't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.5	n/a
	T	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Q	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
can't	P	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.25	n/a	0.4	n/a	0	n/a	n/a	0	n/a	n/a
	D	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	T	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

B.7 Alignment of negation types with motivational states: Additional tables

This section contains the tables which summarize the tables 5.55, 5.56, and 5.57 and which form the basis for the statistical analysis presented in section 5.5. The ANOVAs and t-tests were performed based on the relative frequencies displayed in table B.30.

Table B.29: ***Distribution of motivational states across participants and types (absolute counts).*** *Displayed are the absolute counts of the motivational states the robot was in during the participants' expression of* negative intent interpretations *(*NII*),* negative motivational questions *(*NMQ*), and* prohibition *+* disallowance *(*P+D*) (combined count) for all participants. Abbreviations for motivational states: -: negative,* O*: neutral, +: positive*

	NII				NMQ				P+D			
	-	*O*	*+*	*total*	*-*	*O*	*+*	*total*	*-*	*O*	*+*	*total*
P01	2	0	0	2	0	0	0	0	n/a	n/a	n/a	n/a
P04	22	1	1	24	21	3	0	24	n/a	n/a	n/a	n/a
P05	20	8	10	38	16	13	6	35	n/a	n/a	n/a	n/a
P06	40	14	3	57	7	3	1	11	n/a	n/a	n/a	n/a
P07	35	14	2	51	13	41	25	79	n/a	n/a	n/a	n/a
P08	12	7	0	19	17	22	5	44	n/a	n/a	n/a	n/a
P09	18	11	2	31	9	10	6	25	n/a	n/a	n/a	n/a
P10	1	1	1	3	10	5	2	17	n/a	n/a	n/a	n/a
P11	15	5	1	21	5	2	1	8	n/a	n/a	n/a	n/a
P12	5	3	3	11	1	3	0	4	n/a	n/a	n/a	n/a
P13	14	3	1	18	9	1	2	12	28	13	8	49
P14	0	0	0	0	0	0	0	0	4	0	15	19
P15	26	13	2	41	24	25	9	58	29	5	30	64
P16	12	20	2	34	7	8	1	16	6	2	19	27
P17	11	3	0	14	4	3	0	7	29	0	8	37
P18	21	7	4	32	10	5	1	16	6	0	11	17
P19	17	2	0	19	12	2	0	14	5	2	12	19
P20	15	10	1	26	28	13	3	44	27	4	14	45
P21	4	0	0	4	15	5	1	21	10	14	23	47
P22	0	1	1	2	0	0	0	0	0	1	7	8

Table B.30: ***Distribution of motivational states across participants and types (relative frequencies)***. *Displayed are the relative frequencies (percentages) of the motivational states the robot was in during the participants' expression of* negative intent interpretations *(*NII*),* negative motivational questions *(*NMQ*), and* prohibition + disallowance *(*P+D*) (combined count) for all participants. Abbreviations for motivational states: -: negative,* O*: neutral, +: positive*

	NII				NMQ				P+D			
	-	*O*	+	*num. base*	-	*O*	+	*num. base*	-	*O*	+	*num. base*
P01	100	0	0	2	n/a	n/a	n/a	0	n/a	n/a	n/a	n/a
P04	91.67	4.17	4.17	24	87.5	12.5	0	24	n/a	n/a	n/a	n/a
P05	52.63	21.05	26.32	38	45.71	37.14	17.14	35	n/a	n/a	n/a	n/a
P06	70.18	24.56	5.26	57	63.64	27.27	9.09	11	n/a	n/a	n/a	n/a
P07	68.63	27.45	3.92	51	16.46	51.9	31.65	79	n/a	n/a	n/a	n/a
P08	63.16	36.84	0	19	38.64	50	11.36	44	n/a	n/a	n/a	n/a
P09	58.06	35.48	6.45	31	36	40	24	25	n/a	n/a	n/a	n/a
P10	33.33	33.33	33.33	3	58.82	29.41	11.76	17	n/a	n/a	n/a	n/a
P11	71.43	23.81	4.76	21	62.5	25	12.5	8	n/a	n/a	n/a	n/a
P12	45.45	27.27	27.27	11	25	75	0	4	n/a	n/a	n/a	n/a
P13	77.78	16.67	5.56	18	75	8.33	16.67	12	57.14	26.53	16.33	49
P14	n/a	n/a	n/a	0	n/a	n/a	n/a	0	21.05	0	78.95	19
P15	63.41	31.71	4.88	41	41.38	43.1	15.52	58	45.31	7.81	46.88	64
P16	35.29	58.82	5.88	34	43.75	50	6.25	16	22.22	7.41	70.37	27
P17	78.57	21.43	0	14	57.14	42.86	0	7	78.38	0	21.62	37
P18	65.63	21.88	12.5	32	62.5	31.25	6.25	16	35.29	0	64.71	17
P19	89.47	10.53	0	19	85.71	14.29	0	14	26.32	10.53	63.16	19
P20	57.69	38.46	3.85	26	63.64	29.55	6.82	44	60	8.89	31.11	45
P21	100	0	0	4	71.43	23.81	4.76	21	21.28	29.79	48.94	47
P22	0	50	50	2	n/a	n/a	n/a	0	0	12.5	87.5	8

B.8 Coding scheme for quantitative analysis of negation experiments on a pragmatic level

This section constitutes the coding scheme employed for the 2-coder analysis of negation types and their felicity. The leading theoretical part has been removed from inclusion in this appendix in order not to duplicate the theoretical considerations already elaborated upon in the sections 2.1.2 and 2.1.3. The parts of the coding scheme handed to the 2nd coder as coding manual are marked as such. Note that some proposals brought forward in the scheme such as the consideration of *teaching episodes* have not been implemented by the time of writing of this thesis due to time constraints. We nevertheless left them in the document as they might be useful for future research.

B.8.1 Construction of the Coding Scheme

The purpose of the coding scheme described below is a description of how to determine the negation types for all negative utterances produced by the robot and the participants recorded during the experiments and, furthermore, which productions can be regarded as felicitous or adequate in the given situations.

Against the background of the limitations of speech act theory with regard to the analysis of actual conversations (cf. sections 2.1.2 and 2.1.3) it might not be surprising that these very limitations of SAT's single-utterance approach became apparent right from the start during the first round of coding. The first coding round was conducted by the author who will also be refered to as 1st coder. The initial coding was conducted on the full set of negative utterances taken from the experiments and the resulting types are based on Pea's taxonomy with eventual extension of Pea's taxonomy wherever this was deemed necessary.

Despite the mentioned issues we still think that this mixed qualitative-quantitative approach is useful in order to determine what kind of negation types qua speech acts were produced in experiments of this type. Yet optimally, if the analyst has enough time to do so a complete conversation analytic analysis would be preferable. The usefulness of this mixed approach hinges on the conception of speech acts being extended in order to account of basic conversational phenomena such as adjacency in a way similar to Pea's approach. Without this kind of extension many utterances could not be coded due to them being 2nd part-pairs of adjacency pairs whose theoretical status in Searle's version of SAT is unknown. Thus many of the negation types listed below are, similarily to Pea's types, "types of use" and not 'pure' speech act types à la Searle. They might be rather viewed conversation analytical and speech act theoretical hybrids due to the top-level inclusion of adjacency as defining property on one hand and due to the refusal to accept Searle's short list of possible communicative functions on the other hand. It is unclear if Pea himself would have thought of his types in this way but the fact that his taxonomy uses adjacency as defining criterion on the top-level and his refusal to have the elements of his scheme be classfied as "types of early meaning" leads us to think that he indeed leaned towards a conversation analytical approach with a further leaning to Wittgenstein's "definition via use". In this context it is important to realise that at the time of Pea's publication CA was in its early days.

The underlying idea on how to decide on the felicity of a particular utterance (stage 1) is to let a competent English speaker decide if an utterance was felicitous or not without taking into account the theoretical (and categorical) apparatus as proposed by Searle but by relying instead on the judgemental process that any fluent speaker of a language employs when engaging in conversation herself. As it turned out this ethnomethodological approach has its own problems (see section B.8.3).

The time stamps and durations in the table for each negative utterance were derived from the recorded audio and video, and the robot's log-files. Deechee's utterances are in few cases hardly audible or not audible at all when watching the video recordings as the audio was recorded via the headset worn by the participants and which suppresses external noise. Therefore, instead of relying solely on the audio recording, Deechee's utterances were extracted from the log files of the *knnlanguaging* module - the module that determines during any experimental session content and timing of Deechee's utterances. Using this log data ensured that no robot utterance was missed.

Future work In terms of notation it might be helpful to introduce the concept of a *teaching episode*. Due to the way both experimental scenarios are set up, the interaction and conversation of any session naturally divides into a sequence of *teaching* episodes. Participants always present and explain one object after another with small transition times in between object presentations. A *teaching episode* starts with the participant picking up a particular object in order to present it to Deechee and ends when the object is either put back on the table or if the object is dropped by Deechee. Such a segmentation leads to some parts of the interaction and conversation being outside of any such episode. At the moment the data is not segmented into such episodes, but we think of the introduction of episode boundaries as advantageous for analytical purposes. We hypothesize that the temporal location of episode boundaries have an impact upon how participants react to Deechee's utterances and bodily behaviour, which is encoded in the column *P (re-)acts in accordance with R's behaviour or speech* (see section B.8.4 below).

B.8.2 Selection of sessions for the 2nd coder

On recommendation by one of the psychologists of the research group the percentage of negative utterances to be coded by the 2nd coder was set to 20%. Initially we were planning to select randomly 20% of the 100 experimental sessions. As the number of negative utterances per session and/or time unit vary greatly in between parties for both participants and robot, this method could lead to far less than 20% of the total number of utterances being coded. This situation would occur if sufficient sessions from less-than-average communicative participants were randomly selected. For this reason we decided to select 20% of the negative utterances instead of selecting 20% of sessions. The considerable variation of the number of negative utterances per session is true for both human and robot utterances.

Random selection of sessions for negative robot utterances The following procedure is applied in order to determine which sessions are selected for dual-coder coding of the robot utterances:

Let *TNU* be the total number of utterances across all sessions and participants. Split the (super)set of all sessions into two sets, each of which contains all (sub)sets of a particular condition, rejective condition vs. rejective-prohibitive condition. For each of these sets do:

1. Fill the *big_ bag_ of_ sessions (BBS)* with all sessions of all participants in the selected condition
2. Select randomly a session from *BBS*
3. Determine the number of negative utterances per session of either human or robot
4. Add this session to the *small_ bag_ of_ sessions (SBS)* for dual-coding and remove it from *BBS*

5. If the total number of utterances in the bag equals or exceeds 20% of *TNU* stop, otherwise go back to 1

Note that this procedure bears the following risk: Once a session of a very talkative participant is chosen we could end up with very few session to be coded for the 2nd coder as the chosen session might already constitute a considerable share of the 20% target. Nevertheless the alternative of selecting arbitrary utterances from arbitrary sessions on an utterance-by-utterance basis was abandoned due to the following practical reason: Coding utterances, that are randomly distributed across the whole corpus, would lead to a much higher effort for the 2nd coder as she would have to potentially open dozens of files and skim through each video in order to locate any one particular utterance. In total there are 100 video recordings at 5 minutes each. Due to financial limitations with regard to the payment of the 2nd coder as well as time constraints inherent to any PhD it was decided against this principally favourable but practically very time-consuming method.

Random selection of sessions for negative human utterances In order to select the 20% of the participants' negative utterances the same files were coded as the ones selected by the procedure specified in the last paragraph. As there are more negative utterances produced by the participants than there are produced by the robot, additional files had to be selected to reach the 20% margin: For each scenario all files across all sessions were numbered from 1 to 50. Subsequently a random generator was used to produce a new number upon any new run. The corresponding file was added to the set of files to be coded. This procedure was repeated until the 20% margin was reached.

B.8.3 Coding process

As mentioned in section B.8.1 above, the author and first coder coded all entries for all participants and all sessions during the initial stage when constructing the coding scheme and determining the initial set of negation types to be used by the 2nd coder. The coding process for the 2nd coder is separated into three stages for the two coders in order for the two to be able to work partially in parallel: While the processing of the 1st and 2nd stage by the 2nd coder was under way, the first coder seeked to reduce the number of negation types for stage 3. At that stage there were 24 negation types on part of the participants which was deemed too many. This number has subsequently been reduced to 19 types.

In a potential 4th stage a speech act type analysis could be performed in which the *felicity* or *adequacy* of the robots negative utterances could be determined in terms of Searle-style satisfaction conditions which can probably be mapped to the columns *body behaviour* and *ling. signaling*. The column *P (re-)acts ..* could possibly give some insight into the perlocutionary effect of each utterance - the consequences of a robot utterance onto the participant.

Stage 1 Coding of robot utterances for felicity: The second coder codes the last row (*felicity*) of the table below. Based on her knowledge of English and by virtue of being a fluent speaker of the language she decides if a particular negative utterance is felicitous/adequate, i.e., makes sense in the given situational and conversational context.

Observations: To decide intuitively if a given utterance is adequate in a given situation is often easier said than done when one of the conversation partners has a low communicative competence. One reason for this uncertainty is for example the choice of the coder between looking at the dialogue from an outsider perspective or looking at the dialogue from the perspective of any particular participant. We observed that different participants

reacted very differently to the (linguistic) actions of the robot: some participants were very sensitive to the robot's behaviour and/or speech, others clearly weren't. Therefore, and this is roughly in accordance with satisfaction conditions of illocutionary acts in speech act theory, the 1st coder advised the 2nd coder to primarily judge from an outsider perspective and, if in doubt, to 'side with the robot'. Also parents occasionally ignore what children say, but that does not mean that the child's utterance is conceived of as inadequate in a given situation. The failure in such a situation is a failure on the interactional but not on the speech act level. In other words, we decided in such cases that the blame is primarily assigned to the side of the conversationally adept conversation partner, i.e. the participant, and only secondary to the robot. Notice that in adult-adult conversation where the conversational competence can be assumed to be roughly identical the blame-assignment problem for the observer does not or only marginally occur, as both conversation partners are in possession of the necessary tools for conversational repair. In a case of misunderstanding these tools are put to use which renders the repair process and potential blame assignment transparent to an external observer. In the case of our asymmetric conversation partners, a non-reaction of the participant has no effect on the adequacy of an utterance. There might be good reasons for the participant (or the parent) not to listen to the child or robot but these reasons don't taint the issue if any particular utterance is adequate or not. One may take a different stance here if the history of the whole conversation as well as the personality of the participant is taken into account. But taking the complete history of the conversation into account, if done properly, equates to a full-blown conversation analysis, which is precisely what we could not perform for the given data due to time constraints. The author intends to perform such a full-blown analysis for selected participants. The author also asked every participant to fill out a TIPI form. An analysis of these, yields the possibility of detecting potential correlations between particular rows of the table and

particular personalities. At the current stage of this quantitative-qualitiative analysis this data is not yet available. Furthermore it is unclear to the author, even if this data was available, how particular personality traits ought to influence the decision about the adequacy of the robot's utterances on a per-utterance basis. Taking the participants reaction into account for the purpose of determing adequacy on a speech act level would mean to move towards the realm of perlocutionary acts.

Despite these decisions we still encountered considerable uncertainty in terms of deciding on the felicity of the robot's negative utterances. The author caught himself often going back through the codes and changing some of them after some particular utterance made him change his mind about the 'intuitive' requirements for an utterance to be considered "ok"/adequate. The 2nd coder reported similar problems. Based on this experience the 2nd coder was instructed to write down in ambiguous cases why she decided in a particular way. So even if the agreement of the two coders should turn out to be very low, we will be able to determine why we disagreed on particular occasions by comparing our reasons. Such a comparison might tell us if a 2-coder analysis makes sense at all for this kind of analysis and touches upon some of the questionable foundations on which speech act theory stands which were mentioned above: the very attempt to make a judgment about "felicity" on the level of single utterances/speech acts without (properly) taking the conversational context into account is problematic. The author also observed that the same 'linguistic moves' are repeated by participants within one episode to increase the certainty about what the robot wants. It seems very hard to tease a sequence of such moves apart in order to decide for each separately if it is adequate or not. The principal idea behind separating the coding in stage 1 from the coding in stage 2 is to separate the 'intuitive decision' from a more SAT-style analysis. But as just mentioned this separation is not as clean as we hoped it to be, as the advise to make one's decision on "adequacy" independent of the partici-

pants reaction hinges on the sometimes rather fuzzy boundary between illocutionary and perlocutionary acts. We currently don't see an alternative to this approach apart from performing a complete conversation analysis.

Stage 2 Coding of negative robot utterances for the remaining columns: The second coder codes the remaining columns listed in the table B.31. We expect all columns except columns 7 and 8 (*P (re-)acts in accordance ..*, *negation_ types*) to be unproblematic in terms of inter-coder variance. The set of candidates for each column is described in section B.8.4.

Stage 3 Coding of human negative utterances: The second coder fills in all empty columns listed in table B.32 for each negative utterance listed there, which are all negative utterances produced by the corresponding participant in the session in question. The set of candidates for these columns is described in section B.8.5.

(Stage 4 (future work) Speech Act Theoretic analysis:) If possible, we intend to perform a speech act theoretic analysis. At the moment it is unclear, if this actually is feasible for the speech recorded in our experiments. In the literature very few attempts have ever been made to apply SAT to child and child-directed language. In any case we would expect that some of the negation types listed in section B.8.4 would become obsolete. If feasible, a separate SAT analysis would amount to a reassignment of the values for *felicity* based on formal SAT satisfaction criteria instead of letting two coders decide intuitively on this value. In the best case the SAT criteria would could be mapped to the observations encoded in the columns *body behaviour*, *ling. signaling*, and *P (re-)acts ...* The value for *linguistic signaling* could yield an indiction towards the uptake on the participants side, the *body behaviour* column indicates the psychological state of the robot (sincerity condition

in SAT) and *P (re-)acts..* equates roughly to the perlocutionary effect of an utterance. As already mentioned, for most speech act types the perlocutionary effects are not part of the set of satisfaction conditions which determine felicity. Yet there are exceptions for some types of speech acts such as *making a bet* or *requesting* which have to be conventionally acknowledged by the conversation partner.

B.8.4 Coding Table for Robot Utterances

Columns marked **bold** are given. They were derived from the log files of the module which is responsible for Deechee's speech. Columns for time stamps are omitted for space reasons. Column *negation_word* is omitted as it's not applicable for robot utterances.

Table B.31: *Coding Table for Negative Robot Utterances*

speaker[a]	**utt.**	body behaviour	ling. signaling that utterance was (mis-) understood	P (re-)acts in accordance with R's behaviour or speech	negation type	felicitous
R_utt	**no** **go** **don't** **...**	positive rejective prohibited undecided neutral	signals underst.[b] signals misunderst.(+word)[c] no signal	N/A: see text AB: acts on behaviour AS: acts on speech ASB: acts on both NoA: doesn't react	truth-func. denial neg. agreement mot. dep. denial mot. dep. exclam. neg. imperative persp. dep. denial rejection of offer self-prohibition mot. dep. assertion none	yes no n/a

[a] fixed in advance by coder 1, based on speech log files and 1st round of video analysis
[b] Example: Deechee: *No* P: *no, not a big fan?*
[c] Example: Deechee: *Go* P: *no? ok*

Explanation of the single columns

speaker Specifies the speaker: robot (R) in case of R_utt or participant/human (P) in case of H_utt.

start_of_utt Time relative to the start of the video recording when the utterance started. Note that this time might be inaccurate by around 1 sec.

utterance This field contains the negation word which the robot produced.

body behaviour This field contains a description of the robot's bodily behaviour at the time when it pronounced the given utterance. It is **only applicable to robot utterances**. The following five different types, also listed in table B.31, are possible:

- **positive**: R is smiling and possibly reaching for an object
- **rejective**: R frowns and potentially looks away from object and participant
- **neutral**: the robot is neither smiling nor frowning and neither reaches nor avoids a presented object
- **undecided**: the robot starts to smile and starts to hold its hand out but flinches back again and stops smiling. This "approach and flinch back" move might be repeated in this situation several times
- **prohibited**: the robot is actively prohibited by the participant, that is, the participant restrains the reaching movement of the robot.

linguistic signaling .. This field contains an evaluation that indicates if the participant signaled in some way that he or she understood the robot's utterance. It is **only applicable for robot utterances**. There are three possible values for this field:

- **signals understanding**: P somehow signals that she understood the word. Often this happens by repetition of the word with an assertive intonational contour ($\rightarrow$ neg. type: *neg. agreement*) or an intonation contour of doubt ($\rightarrow$ neg. type: *neg. question*). Sometimes utterances might be deemed signals of understanding that were not mere repetitions of the robot's word, especially if it was witnessed before that the participant in question understood this very utterance. A necessary requirement is that P says something in direct succession ($\Delta t \ll 1s$) to R's utterance, and that this utterance is deemed by the coder to be an affirmative signal that serves to indicate that the utterance was understood.

  If the coder deems a signal to be such a "signal of understanding" (*uptake*) this does not imply that this very signal can not serve other functions as well. Latter functions are outside the scope of this analysis. For example a participant answering Deechee's *no* with a *no?* not only signals that she understood Deechee's *no* but might also signal, at the same time with the very same word by intonating it in a particular way, that she's not sure or convinced that Deechee really means what it just said. The important point is, that the coder thinks that one of possibly many functions of P's utterance is to signal understanding w.r.t. what Deechee just said.

  Example:

  P offers Deechee the heart and Deechee starts to frown.

  R: *No*

  P: *No? Why not? The last time you liked playing with the heart*

- **signals misunderstanding (+ word)**: P somehow signals that he or she misunderstood what the robot said. Upon R uttering a word, P says what she understood in direct succession to R's utterance. The intonational contours are typically identi-

cal to the ones in in the last case: *assertive* or *doubtful*.

Example:

R: *Go*

H: *No? Alright ..*

Here P evidently mis-heard and took the *go* for a *no*. In this case please specify the word for which the 'real' word was mistaken for in brackets after the type.

- **no signal**: participants don't signal understanding or misunderstanding. **This is what is usually the case**.

P (re-)acts in accordance .. The idea behind this column is to see if participants (re-)act on bodily behaviour/gestures and/or speech. When both gestures and speech are in accordance the decision is straight forward. The more interesting case is when gestures and speech are incongruent and when the behaviour leaves ample room for interpretation (typically *neutral* and *undecided* behaviour). Below the term *teaching episode* is used with which the following is meant:

A *teaching episode* starts with the participant picking up a particular object in order to present it to Deechee. The *teaching episode* ends when the object is either put back on the table by the participant or Deechee or if the object is dropped by Deechee. Such a segmentation leads to some parts of the interaction and conversation being outside of any such episode, this is ok. There are five possible values:

- **N/A (not applicable)**: at the time when the negative is uttered participants cannot react to the robot at all or it is highly unlikely that they do so. The situations where this is case are the following:
  - R utters the negative word in between two *teaching episodes*, for example when

P has just put down an object and is looking for the next object to present to R.

- P is constraining R's arm with one hand, as asked for by the instructions, and holds the box with the other. In this situation P's 'freedom to act' is diminished as she follows the experimental instructions.

- **AB (acts on behaviour):** the participant (P) reacts in accordance with the robot's behaviour.

  Example 1:

  R exhibits positive behaviour and says *no*. P tries to put the box into the robots hand (and effectively ignores the *no*)

  Example 2:

  R exhibits rejective behaviour, P starts to put the box down, the robot says *no*

- **AS (acts on speech):** P reacts in accordance with the robot's speech.

  Example 1:

  R exhibits neutral behaviour and P is offering the box to it.
  R: *No*
  P puts the box down, possibly confirming the robots utterance with
  P: *No? Alright*
  Note that a linguistic confirmation is not necessary but is a clear indicator that P actually acts on what R just said.

  Example 2:

The same scenario as the latter but with the robot exhibiting an undecided behaviour

- **ASB (acts on both behaviour and speech)**: P reacts in accordance with both the robot's behaviour speech. This can only be the case if R's behaviour and speech are congruent.

  Example:

  R exhibits rejective behaviour and shortly after says *don't*. P puts the box down after hearing the *don't*, but not before that.

- **NoA (no (re-)action)**: The participant does not react to R's behaviour nor to R's speech.

  Example:

  R exhibits rejective behaviour and says *no*. P ignores both and continues to offer and speak about the box.

negation type Based on a first round of analysis the following negation types for the robot's utterances were derived from the recorded interactions. These types can be split into two groups: adjacent types, and non-adjacent types. *Adjacent* means that the utterance is a linguistic reaction to what the conversation partner said - at this stage of coding the conversation partner is the participant. A response to an answer, for example, is adjacent, adjacent to the answer, whereas answers themselves count in our taxonomy as non-adjacent. A rejection of an offer is non-adjacent, if the offer was performed mainly non-linguistically. (The terminology was most probably coined by con-

versation analysts, see also the very concise article on Wikipedia on adjacency pairs: http://en.wikipedia.org/wiki/Adjacency_pairs).

Adjacent negation types

- **Truth-func. denial**: Truth-functional denials are negative responses to truth-functional assertions or questions. Truth-functional assertions are assertions about state of affairs of the world which can be evaluated objectively. With *objectively* we mean here, that the truth or falsity of the assertion does not depend on the motivational state of the speaker or hearer nor on their perspective.

  Example:

  A: *It's raining outside*

  B: *No (I don't think so)*

  The truth or falsity of "It's raining outside" is independent of A (or B) being happy, grumpy, or sad. It is also independent of the circumstance if A or B can actually see if it rains or not, that is, if any of the two is close to a window or if A is just coming from outside or not. This is what is meant with *independent of their perspective*. As A's *It's raining outside* is a truth-functional assertion, B's *No* counts here as truth-functional denial. Also note that the example is not a classic adjacency pair as *It's raining outside* is not a question. Also note, if A would have asked *Is it raining outside?*, a *No* on part of B would qualify here as *truth-func. denial* as well.

- **Persp. dep. denial**: Perspective dependent denials are negative responses to perspectival (or perspective-dependent) assertions or questions. The truth or falsity of a perspectival assertion depends on either the knowledge (epistemic), the ability, or the physical perspective of an agent.

Example 1 (epistemic):

A: *You know this one here, don't you*?

B: *No, never seen such a thing*

Example 2 (ability):

A: *A big sportsman like you surely can do 50 pushups or not?*

B: *No, not right after a match.*

Example 3 (physical perspective):

A: *Can you see that red bird there on Mr Burns fence?*

B: *No, I can't see over the hedges. You forget that you're quite a bit taller than I am.*

- **Mot. dep. denial**: Motivation-dependent denials are negative responses to motivation-dependent questions or assertions. The answers to motivation-dependent questions depend on the motivational state of the addressee. With motivational states things like likes and dislikes, or wants (and not-wants) are meant.

  Note, that also those questions or assertions are considered motivation-dependent which assume or refer to motivational states without containing motivational or volitional verbs such as *like*, *want*, *fancy*, *feel like* etc. (implicitly motivational).

  Example 1 ('straight-forward' motivational):

  A: *Do you like dogs?*

  B: *No, not really.*

  Example 2 ('volitional'):

A: *Do you want to come along to the cinema tonight?*

B: *No, I don't feeling like going to the movies today.*

Example 3 (implicit)

A: *Will you finally mow the lawn this afternoon?*

B: *No*

Example 4 (implicit)

A: *How about some ice cream?*

B: *No, I'm rather feeling like something savoury at the moment*

Note that in the examples 3 and 4 there is no verb or adjective in A's questions that would qualify as *motivational*. Nonetheless the question involves the motivational state of the addressee by implicitly, that is without lexically referring to said states, asking if the addressee *feels like*, *is up to*, or *willing* to mow the lawn or if she wants some ice cream. Example 3 furthermore gives a hint that the listener must have responded negatively to the very same question in the past, but this is an issue out of our scope.

- **Neg. agreement**: Negative agreement is given if the participant produces a negative utterance of some kind and R agrees with it by uttering a negative as well. Often but not always the negative is a repetition of the negative expression used by the participant. The participant's utterance can have an 'assertional' intonation contour or a question contour. In the latter case it must be a question which already suggests or anticipates an answer as is the case in example 1. Without the question suggesting an answer there would be nothing with which B could agree with.

  Example 1:

A: *So you don't like strawberries then?*

B: *No*

Example 2:

A: *No, evidently you're not very keen on strawberries.*

B: *No.*

Non-adjacent negation types

- **Rejection of offer**: Rejective utterances are very similar to motivation-dependent denial, the main difference being that the latter is adjacent to another utterance of the conversation partner. Rejections are always reactions to non-linguistic offers or proposals of some kind.

  Example:

  A is holding out an apple towards B, effectively offering it to B but not saying a word

  B: *No, thanks! I'm not very much into fruit.*

  It is important to emphasize that the timing of the robots utterances in relation to the human utterances is important to distinguish *rejection of an offer* from *motivation-dependent denial*. The crucial question is: Does the coder deem the utterance to be an answer or another kind of linguistic reaction to a recent utterance of the participant? Only if the coder thinks that the utterance is independent of the participant's utterance(s), it can be a case of rejection.

- **Self-prohibition**: Self-prohibition can only occur in the prohibitive setup. It con-

sists of the repetition of a word which was previously used by the participant in a prohibitive way and/or while physically prohibiting the robot. Often participants counteract self-prohibition by saying things like *No, you can have that.*

Example:

P is holding out a box to the robot

R: *Can't*

P: *Yes, you can hold it*

- **Neg. imperative**: *Negative Imperatives* are similar to *rejection of offers* but don't assume an offer on part of the conversation partner. For an utterance to count as a negative imperative it is necessary though, that the person that is addressed with the imperative is in the process of doing something or just about to do something that is not wanted by the speaker.

  Example:

  A: *And now we're going to put the chicken into the microwave.*

  B: *No! Are you crazy?*

- **Mot. dep. assertion**: Motivation-dependent assertions are utterances other than *rejection of offers* or *neg. imperatives* that are linked to the motivational state to any of the conversation partners. They are in some way a residual class for non-adjacent motivational utterances that are too 'weak', intonationally or also by the mere context in which they are uttered, to count as *neg. imperatives.*

  Example: (mot. dep. assertion in B's 1st utterance)

  A: *I'm going to the cinema tonight! Can't wait to see the new Star Wars!*

B: *I'm still knackered from the weekend. So I guess, I'll give that one a miss.*

Negation types that can be both adjacent and non-adjacent

There is only one type of negation in our taxonomy that does not clearly fit into the adjacent-non-adjacent dichotomy:

- **Mot. dep. exclamation**: Motivation-dependent exclamations can stand in terms of adjacency in isolation. In this case they typically refer to a current event. But they can as well be adjacent and refer to an utterance of the conversation partner to signal disagreement.

  In the non-adjacent case they might be most similar to *mot. dep. assertions* but are typically less articulate and more spontaneous. As opposed to *mot. dep. assertions* they must refer to some current event of some kind which is disagreed with or negative in some other way.
  In the adjacent case they are most similar to *mot. dep. denials* but are not responses to questions or (linguistic) offers but rather express disagreement with an evaluation or assertion of the conversation partner which was not explicitly asked for.

  Example 1 (non-adjacent):

  A accidentally drops a glass of wine onto the white carpet.
  A (or B): *Oh no! I'll never get that stain out.*

  Example 2 (adjacent):

  A and B are watching a football game together but side for opposite teams. Team X scores the first goal of the game 10 minutes before the end of the game and chances are that this will remain the only goal of the game.

A: *Finally! It was about time. That's it then, I guess.*

B: *No way! This game is far from over.*

Example 1 illustrates that the event which triggers the exclamation can be caused be the speaker itself or another agent. This type does not distinguish by whom the triggering event has been caused or if it was caused by any agent at all. Natural events such as a thunderbolt which are not caused by any agent could trigger such an exclamation.

Other types

- **None**: It can happen with those negation words which I qualified earlier as *pragmatic*, such as *go*, *down*, or *done*, that they are sometimes used and/or perceived as negatives, i.e. they may serve for example the function of rejection. This is not always the case though. So if one of these *pragmatic* negatives in a particular situation is none, their type should be qualified as such: *none*. This is only applicable to *pragmatic* negation words, not to regular lexical/ grammatical negatives such as (*no*, *don't*, *can't* etc).

B.8.5 Coding Table for Human Utterances

Columns marked **bold** are given. Each entry was extracted from the files coming out of the prosody analysis. The time stamps were created by hand by the 1st coder when coding each entry. As opposed to the robot utterances the human utterances are tagged with a start and end time to roughly give an idea of where the automatic utterance boundary detection put the boundary. The utterance delimited through these boundaries does not necessarily coincide with what we intuitively think of as an utterance.

Note that there can be more than one negation word per utterance. If this is the case please specify both, separated through a semicolon in the column *negation word*. If you think that the two negation words belong to different negation types, also specify both negation types, separated through semicolon, in the column *negation type*. If you think that both negative words belong to the same negation type, it is ok to specify the type only once in this column.

Explanation of the single columns

speaker Specifies the speaker: robot (R) in case of R_utt or participant/human (P) in case of H_utt.

start_of_utt Time relative to the start of the video recording when the utterance started. Note that this time might be inaccurate by around 1 sec.

end_of_utt Time relative to the start of the video recording when the utterance ended.

salient word This field contains the most salient word of the particular utterance as determined by the prosodic analysis. The most salient word can or cannot be the negation word. This is why this field might contain non-negative words.

Table B.32: *Coding Table for Negative Human Utterances; entries for columns in* ***bold*** *are given, entries for the remaining columns have to be entered by the coder*

speaker	**start_of_utt**	**end_of_utt**	**salient word**	negation word	negation type
	1:17:59	**1:19:20**	**okay**	no	truth-func. denial
	3:49:27	**3:50:10**	**squares**	no	neg. agreement
H_utt	**1:01:49**	**1:03:40**	**don't**	don't	neg. question
	5:17:39	**5:20:30**	**hearts**	no	rejection_of_request
					negating self-prohibition
					truth-func. question
					neg. persp. question
					neg. mot. question
					neg. persp. assertion
					mot. dep. assertion
					truth-func. negation
					prohibition
					disallowance
					neg. promise
					neg. tag question
					neg. intent interpret.
					quoted negation
					mot. dep. exclamation
					neg. imperative

negation word This field has to be filled with the negation word(s) in the utterance by the coder. It might be the same word as in the field "salient word" above or it might differ from it as not all negation words are salient.

negation type Based on a first round of analysis the negation types listed below for participants' utterances were derived from the recorded interactions. These types can be split into two groups: adjacent types, and non-adjacent types. *Adjacent* means that the utterance is a linguistic reaction to what the conversation partner said, in this stage of coding the conversation partner is the robot. A response to an answer, for example, is adjacent, adjacent to the answer, whereas answers themselves count in our taxonomy as

non-adjacent. (The terminology was coined by conversation analysts, though we use the notion of adjacency in a somewhat broader sense. See also the very concise article on Wikipedia on adjacency pairs: http://en.wikipedia.org/wiki/Adjacency_pairs).

In all examples below the negative words which form part of the respective type are underlined like this.

If an example contains square brackets this indicates an overlap of the speech of the robot and the participant, they pronounced the so marked words simultaneously. Square brackets with a number in them indicate pauses, and the number corresponds to the duration of the pause in seconds.

Adjacent negation types

- **Truth-func. denial**: Truth-functional denial is a reaction of the participant to a truth-functional utterance. See also the explanation of the same type in the section on robot utterances.

  Example 1:

  P: *What's this one?*

  R: *Heart*

  P: *No, it's not a heart*

  Example 2:

  P: *Heart!*

  R: *Circle*

  P: *No, bad*

  Example 3:

P: *What about the circular one?*

R: *There*

P: *No, this one's a circle*

- **Neg. agreement**: See the explanation for the same type in the section on robot utterances plus the following extension: If a participant implicitly assumes a negative utterance, as in example 1 below, this also counts as negative agreement. The *either* in example 1 indicates that the participant believes that Deechee doesn't like the object. For all that matters P acts as if Deechee would have explicitly said *no* before.

  Example 1:

  P: *It's not my favourite either, I'll get rid of it.* (Deechee did not say *no* before P uttered this)

  Example 2:

  P: *You want to play*

  R: *No*

  P: *No, ok. Lets try a different one then.*

  Example 3:

  P: *You don't like the heart? No? It's turning away from me, you don't like the heart*

  R: *No*

  P: *No! No, ok.*

- **Neg. question**: Negative questions are very similar to *negative agreement*: they are

adjacent negatives in which the negative word of the conversation partner is repeated. As opposed to *negative agreements* the intonation contour here is one of doubt or surprise. As opposed to *neg. perspective questions* and *neg. mot. questions* (see below) this question type is necessarily adjacent to the utterance of the conversation partner.

Example 1:

P: *Do you [like] squares*

R: *[No*

P: *No?*

Example 2:

P: *Got a circle here*

R: *circle*

P: *Well done, that's right*

R: *No*

P: *No?*

Example 3:

P: *What about the moon? The crescent [there?*

R: *[No*

P: *No?*

- **Rejection_of_request**: As the name indicates, a rejection of a request is given if Deechee asks the participant for something (or the participants interpret Deechee's utterance in this way) and P rejects linguistically to comply with Deechee's request.

As opposed to *rejection_ of_ offers* (see section on the classification of robot utterances), *rejections of requests* are adjacent to a linguistic request of the conversation partner.

Example 1:

R: *Moon*

P: *No, you've had the moon already*

Example 2:

R: *Square*

P: *You no.. no I'm not gonna show you the squares any more*

Example 3:

R: *Moon*

P: *No, we've had the moon already*

- **Negation of self-prohibition**: Negations of self-prohibitions only occur in the prohibitive scenario. Deechee utters a prohibitive negative, that is, a negative which was previously used by the participant to prohibit Deechee from touching a box. In this situation participants sometimes interpret Deechee's utterance as a form of self-prohibition and counteract using this type of negation.

  Example 1:

  P presents and speaks about the heart box. R goes for it, but flinches back.

  R: *No*

  P: *No, you're allowed to touch this, it's ok*

  Example 2:

P speaks about the hearts, R smiles and reaches for it and P hands the box to R

R: *No*

P: *No, you can hold it, I don't mind*

Example 3:

R is holding the moon box. P used *never* in previous sessions to explain Deechee which boxes were forbidden.

R: *Never*

P: *No, no, good. This is for you*

Non-Adjacent Negation Types

- **Truth-func. question**: Truth-functional questions, as the name indicates, are questions that refer to or ask for some state of affairs being or not being the case. The way they were typically used by participants in case of them being captured in the coding table was to contain suggestions to possible answers. Open truth-functional question, that is questions which do not already suggest an answer or a set of possible answers, albeit being the norm within the experiments are typically not listed in our table due to a lack of negative words.

  Example 1:

  P: *Is that a square? Yeah, no?*

  Example 2:

  P: *Is that a heart? No?*

- **Truth-func. negation**: Truth-functional negation is supposed to capture all kinds of truth-functional negation which are not *truth-functional denials*. Truth-func. negation is in this sense a residual class that captures all non-adjacent truth-functional utterances, be they negative assertions, suggestions, speculations, or guesses about state of affairs, which are in essence truth-functional. Also negative normative assertions such as the one in example 2 below count as a member of this class. Normative assertions are assertions about rules, laws or general practices in society such as "Thou must not kill" (law), "When driving a car one must stop in front of a red traffic lights" (rule), or referring to social practices in Italy, "When you greet somebody it is common to give the person two kisses, one on each cheek" (social practice).

  Example 1:

  P: *My heart beats. Have you got one?* [1.5s] *No robot heart maybe? Maybe not.*

  Example 2:

  In the context of explaining round traffic signs:
  P: *They will tell you 30, which means you mustn't go any faster than 30.*

  Example 3:

  P: *Which one didn't we look at? We didn't look at the moon.*

- **Neg. persp. question**: Negative perspective questions, together with positive perspective questions, are the counterpart to *perspective dependent denial* on the side of the robot utterances. As is the case with the latter, they encompass questions, where the truth of the answer depends on either the knowledge (epistemic), the ability, the physical perspective of the agent or any other state of affairs which can be

only judged by the agent that is addressed. For example, in example 2 below only the addressee can decide whether it is hungry or not. These questions either contain a lexical negative such as *no* or a grammatical one such as *don't*.

Example 1 (epistemic):

P: *Do you remember the moon? Don't you remember the moon?*

Example 2 (physical perspective):

P: *Can you see the squares? No? Ok*

Example 3 ("biological perspective"):

P is speaking about lollipops

P: *No? You're not feeling very hungry today?*

- **Neg. persp. assertion**: Negative perspective assertions, together with positive perspective assertions, can be found as counterpart to *perspective dependent denial* on the side of the robot utterances. All remarks on the dependencies of truth values mentioned under *neg. persp. questions* apply here as well. Furthermore perspectival assertions are captured here which are not about some perspectival aspect of the addressee but about such an aspect of the speaker (see example 1). Sometimes it's very hard to distinguish *neg. persp. assertions* from *neg. persp. questions* as for example in example 2.

  Example 1 (epistemic, regarding the speaker herself):

  P: *Do you like the one with the squares? I can't remember*

  Example 2 (epistemic, regarding the addressee)

P: *You don't remember this one?*

Example 3 (physical ability)

P tries to balance a box on Deechee's hand

P: *No, I don't think you can hold it.*

Example 4 (physical perspective)

P: *Can you see it? You can't? Or you're looking ..*

Example 5 (other ability)

P: *Can you say moon? No*

- **Neg. mot. question**: Negative motivational questions are questions that contain a lexical or grammatical negative. In the extreme case they consist of nothing else than this very negative, which has the intonational contour of a question. They may refer to the motivational state of the addressee directly (example 2). But they may also refer to stances or preferences (example 1) or intentional actions (example 3) that are indirectly linked to motivational states. In the direct case the question contains motivational or volitional verbs or constructions such as *want*, *like*, *feeling like*, *being keen on* etc. In the indirect case they do not contain such 'motivational markers' but clearly refer to the preferences of the addressee or her willingness to perform a certain action based on her current motivation. As it happens, example 3 contains the volitional word *want*, but a pragmatically equivalent question in the given context might be "Are you going to hold it? No?" which does not contain any such markers.

  Example 1:

P: *They're pretty. Don't you think hearts are pr.. . I think hearts are very pretty.*

Example 2:

P: *You wanna look at the circles again. Do you not like the heart?*

Example 3:

P: *Do you want to hold it? No?*

- **Neg. intent interpret.**: Negative intent interpretations are assertions in which the participant interprets Deechee's intentional or motivational state utilizing lexical and/or grammatical negatives. Typically the semantics of these expressions is negative as well, i.e. the participants expresses that she thinks that Deechee does "not want" or "not like" either a particular object or does "not want" or "not like" to perform a particular action such as holding the box. Neg. intent interpret. are in some way a sub-type of *mot. dep. assertions* (see below). Whereas *mot. dep. assertions* can refer to present, past, or future motivational states of speaker or addressee, *neg. intent interpret.* refer to the motivational states of the addressee only and only of his or her states right here and now. They are thought to have a special importance in early language acquisition in that toddlers might learn what we call here motivational words and their meaning by way of caretakers interpreting the toddler's emotional states or intents linguistically.

  At times it can be hard to distinguish between *neg. mot. questions* and *neg. intent interpretations* as the main difference between the two types is the fact if the utterance is a proper question or not.

  Example 1:

P: *No you don't like circles do you like triangles* (no transcription error here, the participant indeed merged two expected *you*'s into one)

Example 2:

P: *You don't want to hold the box. [1s] No*

Example 3:

P: *Do you like the circles box?*

R: *circles*

[1.5s]

P: *No? Ok*

- **Mot. dep. assertion**: Motivation-dependent assertions are all assertions that refer directly (example 1) or indirectly to the motivational states in the present (examples 3), past (example 1) or future of the speaker, or the addressee (example 2), which are not *negative intent interpretations*. This type is in this regard another residual class.

  Example 2 is a borderline case as it is questionable how tightly a personal judgment about a mishap (reading variant 1) is linked to motivational states. Another way of interpreting this utterance (reading variant 2, socio-linguistic) is the following: The purpose of the utterance is to soothe the potential fear of the conversation partner, a fear that might be directed towards the socially dominant teacher, in case of the teacher being angered by the child/robot. Teachers have, by virtue of their social status, the power to hand out punishments. If one accepts this reading variant, there are expected emotional states with both the speaker and the addressee: expected fear on the part of the student (S) based on his expectation of anger on the part

of the teacher (T). T expects S to (potentially) display fear because T expects S to (potentially) expect T to become angry, because S dropped the box. Therefore T counteracts the expected fear by issuing the utterance in order to convey to S that T will not become angry because of what happened. As the term "motivational" here also captures emotional states, a link to motivation would be given under this reading.

Example 1:

P: *And I think the square you didn't like*

Example 2:

P: *Don't worry, not serious (when R drops the box)*

Example 3:

P: *Squares* [1.5s] *I don't like squares, I think they are boring.*

- **Prohibition**: (Linguistic) prohibition only occurs in the *rejective + prohibitive* scenario. It encompasses occurrences of negation whose function is to keep Deechee from touching forbidden objects. Sometimes such an utterance taken in isolation does not indicate that its function is prohibitive, as for example in example 2, which looks rather like a *truth-func. negation*. But, in context, when looking at the video recording, it becomes clear that the utterance is used as prohibition. The prohibitive utterance can or cannot be accompanied by the participant physically restraining Deechee's arm movement.

  Example 1:

  P: *No, no, you're not allowed to touch* (no physical restraint on the part

of the participant)

Example 2:

P: *No, you're not holding it, but you can look at them* (no physical restraint)

Example 3:

P: *No* (P pushes Deechee's arm away)

- **Disallowance**: Disallowance is similar to prohibition but rather captures those utterances that express general (negative) rules. In this sense disallowance utterances are more detached from the here and now of the interaction than prohibitive utterances. Whereas prohibitive utterances are always triggered by a current action on part of the robot, disallowances can or cannot accompany such an action. It can be tricky to clearly distinguish the two types from each other and possibly there is no clear-cut boundary. But there seems to be an important difference between *stating a (negative) rule* on one hand and *uttering a prohibition* with the purpose of stopping an agent's actions at that very moment on the other. The question, that the coder has to ask herself is: Is this utterance meant to act upon Deechee immediately or is it rather the expression of a (general) rule. I observed that both can happen more or less at the same time by uttering a prohibition followed by the statement of a negative rule.

  Example 1:

  P: *You can't have this one* (Deechee is neither being restrained when this is uttered nor shortly afterwards)

Example 2:

P: *You can't touch the moon* (Deechee is not even trying to touch the moon at the time of utterance)

Example 3:

P: *You're not allowed to touch the circles* (This utterance was uttered at times when P was restraining Deechee's arm as well as when not restraining it)

- **Neg. promise**: Negative promises are those negative utterances, in which participants commit themselves not to do certain things (any more) *in the future*. Often said commitment is triggered by a negative reaction of Deechee. "Promise" is actually a slightly too strong term as our category is supposed to capture all kinds of future commitments by the participants - also commitments whose force is weaker than that of a promise. In the examples 1 and 2 what is actually said is the following: "We won't play with X any more". As the participants are in our setup the ones who decide what is played with, this utterance amounts to "In the future I won't pick up X any more".

  Example 1:

  P: *You don't like the circles, no? Ok, we <u>won't</u> play with the circles.*

  Example 2:

  P: *Alright, I'm <u>not</u> gonna force you.*

  Example 3:

P: *Ok, we won't play with that one then.*

- **Neg. tag question**: Negative tag questions are negative grammatical constructions that are attached to the end of the utterance. They consist of the negated auxiliary verb of the main clause, if there is one, plus a personal pronoun (see example 2: *can* [main clause] → *can't you* [tag question]). As can be seen in the examples 1 and 3 the main clause does not always contain the non-negated form of the negated auxiliary verb in the tag question, but putting it there wouldn't make a semantic difference to the utterance (ex. 1: Oh you do like that, don't you, ex. 2: But you did like the circles box, didn't you). Tag questions, negated or non-negated, are not proper questions but are attached to assertions. The negation is purely grammatical and, as far as we know, does not serve any of the functions that the other negatives in our taxonomy serve. Yet as they are distinct grammatical constructs, they are very easy to spot.

  Example 1:

  P: *Oh you like that, don't you?*

  Example 2:

  P: *You can say square for me, can't you?*

  Example 3:

  P: *But you liked the circles box, didn't you?*

Negation types that can be both adjacent and non-adjacent

- **Quoted negation**: In the case of quoted negation, the negative part of the utterance belongs to a part of reported speech, which, if written down, could be quoted or would constitute indirect speech. The speech reported can either stem from the participant herself or from Deechee.

  Example 1:

  P: *I said 'no'. Not this* (uttered in a prohibitive situation)

  Example 2:

  P: *No, you don't like it. You said you didn't like the squares.*

  Example 3:

  P: *What you're saying Deechee? No? Ok*

- **Mot. dep. exclamation**: See explanation to the same type in the section on the coding of robot utterances.

  Example 1:

  P: *Clever boy, I didn't even need to say the name!*

  Example 2:

  P: *Oh we go back to the crescent moon then I think you quite .. oh dear, oh no!*

- **Neg. imperative**: Negative imperatives are a residual class for all those imperatives which are neither *prohibitions* nor *disallowances*. The two latter types cover all imperative negatives which are linked to the prohibition task, set out in the experimental instruction, that is, that Deechee must never touch the forbidden objects.

Example 2 below was not conceived of as a form of disallowance as it was judged to be more general than utterances that are tightly linked to the prohibition task. In the case of this example this judgment is supported by the fact that the participant used different, and more specific prohibitive utterances, which refered specifically to the particular situation at hand (the prohibition task), before uttering this negative.

Example 1:

P: *Oh you're holding that very nicely. Don't throw it away*

Example 2:

P: *You can't have it!* (uttered while P is

P: *It's no good, it's no good putting a face like that* restraining Deechee)

Example 3:

P: *No, don't say 'done'*

END OF SECTION FOR 2ND CODER

B.8.6 Fused Negation Types

The following types were fused in order to reduce the total number of negation types observed amongst the human utterances.

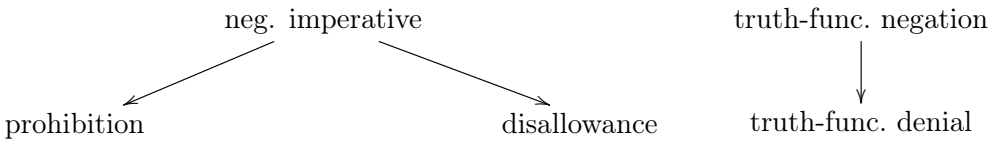
Neg. epistemic assert.
Neg. persp. assert. } Neg. persp. assert.

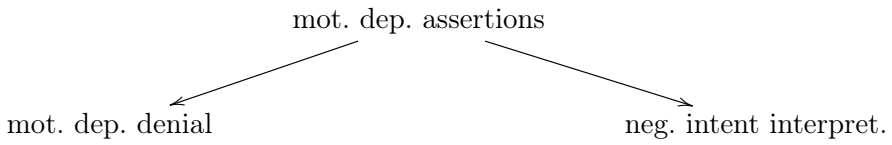
Neg. epistemic question
Neg. persp. question } Neg. persp. question

Apart from an overall reduction of the number of negation types the indicated types were fused into the indicated types because the resulting type subsequently matches the type *persp. dep. denial* on the robot's side as first pair parts.

B.8.7 Super-/Sub-Types

The following types stand in a super-/sub-type relationship to each other, that is, the ones listed at the end of an arrow are more specialized than their super-type. All utterances that are instances of the sub-type are therefore also instances of the super-type.





For the sake of simplicity of the coding scheme one might argue that it would be advantageous to eliminate these sub-types. Yet in order to keep the human negation types as synchronised with the robot negation types as possible, we decided to maintain *truth-func. denial*. Moreover this is also one of the negation types listed in Pea's taxonomy. Another argument for maintaining this type is the circumstance that it indicates a genuine interaction between participant and robot, whereas the super-type *truth-func. negation* is more detached from the actual interaction both in terms of adjacency and in terms of the topic. The same argument can be made for the relationship between *mot. dep. assertions*, *mot. dep. denial*, and *neg. intent interpretations*. Furthermore *neg. intent interpretations* are linked to one of our hypotheses and are thought to play a central role in the acquisitions of emotional/affective words.

With regards to the three types *neg. imperative*, *prohibition*, and *disallowance*, we decided not to eliminate the two subtypes, as *prohibition* is tightly linked with the prohibitive task and shows that actual linguistic prohibition took place, which in general cannot be said about *neg. imperatives*.

B.8.8 Problematic Columns

P (re-)acts in accordance .. [Robot utt.] This column might be the most problematic one in terms of inter-coder agreement. Initially we thought it might be beneficial to restrict the situational context considered by the coders to a very tight time window around the occurrence of the utterance in order to see the 'differential effect' of the negative word, i.e. the change in the participants' behaviour caused by the robot uttering the word.

Consider the following example, where the coder has to decide if the participant is reacting to R's behaviour or not. The decision is one between "no (re-)action" (*NoA*) vs. "acts on behaviour" (*AB*). Consider the situation where the participant holds a box in front of the robot and expects it to take it. The presentation of the box typically lasts a few seconds. Let's say P starts at time x to offer the box to R. R exhibits for $x + 5s$ an undecided behaviour and switches afterwards to a positive behaviour. After $x + 6s$ R says *no*, that is, it produces a negative utterance which is incongruent with its positive behaviour. P tries to put the box into R's hand. If now a tight time window of 2s around the utterance at $x + 6s$ is applied, P's behaviour does not change within this window - P is during all this time holding out the box and trying to get R to hold it. This would result in an entry of *NoA* in this column, as P's behaviour did not change within the time window around the utterance. If we extended the time window to $6s$ though, the entry would change to *AB* because P initiated the movement at the very beginning of the window.

For one single participant and one single session one might be able to find a good value for the time window, but certainly it is close to impossible to find such a window for all participants across all sessions. Some participants might be in a hurry in one session and are therefore less patient with the robot. We observed that some participant react very quickly to R's gestures and speech (decision time $\ll 1s$), whereas others spend ages with 'negotiating' (see next paragraph).

Another problem that a coder faces when deciding which entry to choose for this column is caused by the fact that the willingness to react to the robots gestures and or speech seems, at least partially, to depend upon the relative time from the start of the teaching episode. If P just picked up an object she typically does not react immediately upon initiating this action to the robot utterances and/or behaviour. What is 'typical' in terms of the participants 'reactivity' has to be determined from earlier sessions and the

participant's previous reaction within the same session.

Most participants spent some time with the same box to make sure that R means what it displays/says. For these participants an immediate reaction after initiating the presentation of the box is atypical. Yet there are some participants that display a very high sensitivity to R's behaviour and speech and will probably react to a *no* from Deechee pretty much from the very beginning of the teaching episode. The majority of participants, even if in principle sensitive to R's speech and behaviour do not react immediately upon the start of a teaching episode to R's behaviour. This observation hints to another limit of a Searle-type speech act theory and any other theory that rips utterances out of the conversational context and treats them in isolation: Some pragmaticists take the stance that the meaning of communicative acts is typically negotiated in a conversation (process view as opposed to a 'structure view'). Any negotiation process obviously takes a certain time. Our suspicion is that this is what we have observed in the interactions. Conversational analysts would argue that the relative position of an utterances within the conversation is crucial to understand its overall function within the conversation. Yet our very coding here is not purely based upon CA but rather a mixed quantivative-qualitative approach. And just by the virtue of each negative utterances having a separate entry in the table as is needed for a quantitative approach we treat them, at least tendentially, as independent from each other. Even if the coders do not treat them independently from each other and try to account for their interdependency, any subsequent statistic will effectively treat them as independent.

For these reasons we decided to abandon the 'tight time window' and left the decision to the coder which parts of the participant's behaviours she would deem relevant for making the decision no matter how long before the time of the utterance the behaviour was initiated. If for example the robot is holding out the hand for 15s, the participants puts

the box into it's hand at x+8s, and R says "no" at x+10s, we would still code this as *AB* despite the relevant action having been initiated long before the utterance.

Appendix C

Conversation Analytical Transciption Glossary

The following transcription symbols are used within this thesis and accord with common conversational analytic standards as described in Hutchby and Wooffitt (1999) and ten Have (2007).

Notice that all punctuation marks refer to prosodic speech characteristics, not to grammatical units.

(0.5) or (.5)	Length of a gap in seconds (before decimal point) and 10ths of secods (after decimal point)
(.)	Pause of less than 0.2 seconds
=	'Latching' between utterances: equal signs at end of one utterances and at beginning of a subsequent utterance indicates no 'gap' between the two (parts of) utterances
.hh	Dot before 'h' indicates that speaker is breating in - the more 'h's, the longer the drawing of breath

hh	Without dot the 'h' indicates that the speaker is breathing out
.pt	indicates a 'lip-smack'
[ ]	An opening square bracket that circumbraces two subsequent lines or two vertically aligned square brackets indicate the start of overlapping talk. The closing square bracket indicates the end of the overlapping talk.
(( ))	Double opening and closing brackets indicate non-verbal activity or other comments of the transcriber
-	the dash indicates the abrupt ending of a word or a sound, as it for example happens in self-repair when changing the word
:	The preceding sound or letter has been streched by the speaker if it is followed by a colon. The more colons, the longer the prolongation.
!	The exclamation mark indicates prosodic emphasis at the end of an utterance
.	The period indicates a stopping fall in tone at the end of an utterance, which does not necessarily coincide with the end of a grammatical sentence
?	Question marks indicate a utterance-final rising intonation as is typical but not exclusive to the end of intonation contours of questions.
↑ ↓	Vertical arrows pointing upwards and downwards indicate a rising or falling intonation contour and can be applied within an utterance

(utt)	Transcriptions that occur in single opening and closing parentheses indicate uncertainty on part of the the transcriber, when an utterance is barley audible or otherwise distorted.
( )	Content-less opening and closing single parentheses indicate that the speaker did utter something which was unintelligible to the transcriber.
CAPITALS	Capitalized transcription were perceived as noticeably louder by the transcriber as compared to the previous and following utterances.
° °	Degree signs at the beginning and end of an utterance indicate that this utterance is noticably quieter as compared to the surrounding utterances.
> <	An utterance that is bracketed by a *More than* and *less than* sign was produced quicker ("sped up") as compared to the speaker's average speed of talk.

Appendix D

Contents of the DVD

D.1 <DVD_ROOT>/data

The data directory contains all files related to the audio processing and symbol grounding. The original audio files were compressed into the mp3-format in order to make the entirety of them fit onto the DVD.

D.2 <DVD_ROOT>/forms

This directory contains the consent forms containing the written instructions that were given to participants prior to the experiment proper.

D.3 <DVD_ROOT>/software

The software directory contains all software that was written related to the work presented within this thesis. The subdirectory `scripts_data_proc` contains all the scripts, that were used within the analysis of the data. The `italk` subdirectory contains a snapshot of the

ITALK svn versioning system roughly at the time when the experiments ended. Notice that the software is at the time of writing already out-of-date due to the fast development cycle of iCub-related software. In order to run our modules on the iCub, they certainly will have to be adapted to the particular version of the iCub. We also expect that most time constants, that are extremely important if realistic behaviour as seen in the videos shall be achieved, have to be fine-tuned again.

D.4 `<DVD_ROOT>/videos`

The `videos` subdirectory contains all the experimental video files for both experiments and some demo videos that were made for presentations. The videos are arranged by session. *P01* to *P12* are participants from the rejection experiment, and *P13* to *P22* are participants from the prohibition experiment. All videos were reduced in image size and compressed in order make them fit onto the DVD and are therefore of mediocre quality. The accumulated size of the original HD video files exceeded 100 GB and can therefore not be delivered with the thesis.

D.4.1 Selected scenes

In the following are pointers to selected scenes that show some of the robot's behaviours and noteworthy 'highlights' of the interaction, that we alluded to from within the main section of the thesis.

Head shakes

Participants interpreting the object avoidance behaviour as head shakes — `<DVD>/videos/session4/P15-280312.mp4 5:10`

Drops - intentional or else

iCub dropping/throwing away an object	<DVD>/videos/session2/P05-011211.mp4 5:35
	<DVD>/videos/session5/P06-151211.mp4 0:15
	<DVD>/videos/session1/P07-191211.mp4 0:18
	<DVD>/videos/session2/P08-040112.mp4 4:25
P07 soothing Deechee after accidental drop	<DVD>/videos/session2/P07-040112.mp4 4:53
P07 interpreting drop as intentional	<DVD>/videos/session1/P07-191211.mp4 0:33
iCub holding an object for a while and finally, unintentionally dropping it	<DVD>/videos/session1/P07-191211.mp4 5:22
Giving back a box	
iCub giving back a box	<DVD>/videos/session3/P05-021211.mp4 0:58
iCub giving back a box	<DVD>/videos/session4/P05-051211.mp4 3:27
	<DVD>/videos/session4/P06-141211.mp4 0:28
Negative intent interpret.	
	<DVD>/videos/session4/P09-160112.mp4 5:04

Bibliography

ARToolKit: 2012, `http://www.hitl.washington.edu/artoolkit/`, last visited 4th of October.

Austin, J. L.: 1975, *How to Do Things with Words*, Harvard University Press.

Baillie, J.-C. and Ganascia, J.-G.: 2000, Action categorization from video sequences, *in* W. Horn (ed.), *ECAI 2000 Proceedings*, number 54 in *Frontiers in Artificial Intelligence and Applications*, IOS Press, pp. 643–647.

Baker, N. and Nelson, K.: 1984, Recasting and related conversational techniques for triggering syntactic advances by young children, *First Language* **5**(1), 3–21.

Baldwin, D. and Meyer, M.: 2007, How Inherently Social is Language?, *in* E. Hoff and M. Shatz (eds), *Blackwell Handbook of Language Development*, Blackwell Publishing, chapter 5, pp. 87–106.

Barsalou, L. W.: 1999, Perceptual symbol systems, *Behavioral and Brain Sciences* **22**(4), 577–660.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L. and Volterra, V.: 1977, From gesture to first word: On cognitive and social prerequisites, *in* M. Lewis and L. A. Rosenblum

(eds), *Interaction, Conversation and the Development of Language*, Wiley, New York, pp. 247–307.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L. and Volterra, V. (eds): 1979, *The Emergence of Symbols: Cognition and Communication in Infancy*, Academic Press, New York.

Bates, E., Bretherton, I., Shore, C. and McNew, S.: 1983, Names, gestures and objects: Symbolization in infancy and aphasia, *in* K. E. Nelson (ed.), *Children's Language*, Vol. 4, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp. 59–125.

Bates, E., Camaioni, L. and Volterra, V.: 1979, The acquisition of performatives prior to speech, *in* E. Ochs and B. B. Schieffelin (eds), *Developmental Pragmatics*, Academic Press, New York, pp. 111–129.

Bateson, M. C.: 1979, The epigenesis of conversational interaction: A personal account of research development, *in* M. Bullowa (ed.), *Before Speech: The Beginning of Interpersonal Communication*, Cambridge University Press, New York, pp. 63–77.

Beebe, B. and Stern, D.: 1977, Engagement-disengagement and early object experience, *in* N. Freedmount and S. Grand (eds), *Communicative Structures and Psychic Structures: A Psychoanalytic Interpretation of Communication*, Springer, New York, pp. 35–55.

Bloom, L.: 1970, *Language Development: Form and Function in Emerging Grammars*, M.I.T. Press.

Bloom, P.: 1998, Theories of artefact categorization, *Cognition* **66**(1), 87–93.

Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C. L. and Lee, D.: 2006, The art of designing robot faces - dimensions for human-robot interaction, *in* M. A. Goodrich,

A. C. Schultz and D. J. Bruemmer (eds), *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot interaction*, ACM, pp. 331–332.

Brazelton, T. B., Koslowski, B. and Main, M.: 1974, The origins of reciprocity: The early mother-infant interaction, *in* M. Lewis and L. A. Rosemblum (eds), *The Effect of the Infant on its Caregiver*, John Wiley and Sons, New York, pp. 49–76.

Bruner, J. S.: 1973, Organization of early skilled action, *Child Development* **44**, 1–11.

Bruner, J. S.: 1975a, From communication to language - A psychological perspective, *Cognition* **3**(3), 255–287.

Bruner, J. S.: 1975b, The ontogenesis of speech acts, *Journal of Child Language* **3**, 255–287.

Bruner, J. S.: 1982, The origin of action and the nature of adult-infant transaction, *in* E. Z. Tronick (ed.), *Social Interchange in Infancy: Affect, Cognition and Communication*, University Park Press, Baltimore, pp. 23–36.

Bruner, J. S.: 1983, *Child's Talk: Learning to Use the Language*, Oxford University Press, Oxford.

Bruner, J. S.: 1998, Routes to reference, *Pragmatics & Cognition* **6**(1-2), 209–227.

Bullowa, M. (ed.): 1979, *Before Speech: The Beginning of Interpersonal Communication*, Cambridge University Press, New York.

Cangelosi, A.: 1999, Modeling the evolution of communication: From stimulus associations to grounded symbolic associations, *in* D. Floreano, J.-D. Nicoud and F. Mondana (eds), *Advances in Artificial Life - 5th European Conference, ECAL'99 Lausanne, Switzerland, September 13-17, 1999 Proceedings*, Vol. 1674 of *Lecture Notes in Computer Science*, Springer, pp. 654–663.

Cangelosi, A.: 2001, Evolution of communication and language using signals, symbols, and words, *IEEE-EC* **5**, 93–101.

Carpenter, M., Nagell, K. and Tomasello, M.: 1998, Social cognition, joint attention, and communicative competence from 9 to 15 months of age, *Monographs of the Society for Research in Child Development* **63**, 1–176.

Carter, A. L.: 1975, The transformation of sensorimotor morphemes into words: a case study of the development of 'more' and 'mine', *Journal of Child Language* **2**, 233–250.

Caselli, M. C.: 1990, Communicative gestures and first words, *in* V. Volterra and C. J. Erting (eds), *From Gesture to Language in Hearing and Deaf Children*, Springer-Verlag, Berlin, pp. 56–67.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L. and Weir, J.: 1995, A cross-linguistic study of early lexical development, *Cognitive Development* **10**(2), 159–199.

Choi, S.: 1988, The semantic development of negation: a cross-linguistic longitudinal study, *Journal of Child Language* **15**(3), 517–531.

Clark, E. V.: 2009, *First Language Acquisition*, Cambridge University Press.

Cohen, J.: 1960, A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement* **20**(1), 37–46.

Collis, G. M.: 1977, Visual co-orientation and maternal speech, *in* H. R. Schaffer (ed.), *Studies in Mother-Infant Interaction*, Academic Press, London, pp. 355–375.

Common Environmental Noise Levels: 2013, `http://www.chchearing.org/noise-`

center-home/facts-noise/common-environmental-noise-levels, last visited 1st of September 2013.

Companion Website for: Word Frequencies in Written and Spoken English by Geoffrey Leech et. al: 2001, http://ucrel.lancs.ac.uk/bncfreq/, last visited 14th of May 2013.

Cost, S. and Salzberg, S.: 1993, A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features, *Machine Learning* **10**(1), 57–78.

Daelemans, W. and van den Bosch, A.: 2005, *Memory-Based Language Processing*, Cambridge University Press.

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A. and Blow, M.: 2009, KASPAR - a minimally expressive humanoid robot for human-robot interaction research, *Applied Bionics and Biomechanics* **6**(3 & 4), 369–397.

Di Eugenio, B.: 2000, On the usage of Kappa to evaluate agreement on coding tasks, *Proceeding of LREC*, Vol. 1, pp. 441–444.

Dominey, P. and Boucher, J.: 2005, Learning to talk about events from narrated video in a construction grammar framework, *Artificial Intelligence* **167**, 31–61.

Dore, J.: 1974a, Communicative Intentions and Speech Acts in Language Development, *Technical report*, Baruch College, City University of New York.

Dore, J.: 1974b, A pragmatic description of early language development, *Journal of Psycholinguistic Research* **3**(4), 343–350.

Dore, J.: 1975, Holophrases, speech acts and language universals, *Journal of Child Language* **2**(1), 21–40.

Dore, J.: 1983, Feeling, form and intention in the baby's transition to language, *in* R. M. Golinkoff (ed.), *The Transition from Prelinguistic to Linguistic Communication*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp. 167–190.

Ely, R. and Gleason, J. B.: 1995, Socialization across contexts, *in* P. Fletcher and B. MacWhinney (eds), *The Handbook of Child Language*, Basil Blackwell, Oxford, pp. 251–270.

Ervin-Tripp, S.: 1977, Wait for me, roller skate!, *in* S. Ervin-Tripp and C. Mitchell-Kernan (eds), *Child Discourse*, Academic Press, New York, pp. 391–414.

Erwin-Tripp, S. and Miller, W.: 1977, Early discourse: Some questions about questions, *in* M. Lewis and L. A. Rosenblum (eds), *Interaction, Conversation and the Development of Language*, Wiley, New York, pp. 9–25.

faceAPI: 2013, `http://www.seeingmachines.com/product/faceapi/`, last visited 1st of September.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. and Pethick, S. J.: 1994, Variability in Early Communicative Development, *Monographs of the Society for Research in Child Development* **59**(1), i–185.

Ferreirós, J.: 2001, The Road to Modern Logic - An Interpretation, *The Bulletin of Symbolic Logic* **7**(4), 441–484.

Filipi, A.: 2001, *The organisation of pointing sequences in parent-toddler Interaction*, Unpublished doctoral dissertation, Monash University.

Filipi, A.: 2002, Before speech: The design of early repair in pointing sequences. Pa-

per presented at the International Conference of Conversation Analysis, Copenhagen, Denmark.

Filipi, A.: 2007, A toddler's treatment of *mm* and *mh mm* in talk with a parent, *Australian Review of Applied Linguistics, Special Thematic Issue: Language as Action: Australian Studies in Conversation Analysis* **30**(3), 33.1–33.17.

Filipi, A.: 2009, *Toddler and parent interaction: the organization of gaze, pointing, and vocalisation*, John Benjamins Publishing Company, Amsterdam, The Netherlands.

Fischer, K., Foth, K. and Rohlfing, K.: 2011, Is Talking to a Simulated Robot like Talking to a Child?, *Proceedings of ICDL-EpiRob 2011*, Vol. 2, pp. 1–6.

Fischer, K., Lohan, K. and Foth, K.: 2012, Levels of embodiment: Linguistic analyses of factors influencing HRI, *Proceeding of the 7th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 463–470.

Fogel, A.: 1977, Temporal organization in mother-infant face-to-face interaction, *in* H. R. Schaffer (ed.), *Studies in Mother-Infant Interaction*, Academic Press, New York, pp. 119–152.

Förster, F., Nehaniv, C. and Saunders, J.: 2011, Robots that say 'no', *in* G. Kampis, I. Karsai and E. Szathmáry (eds), *Advances in Artificial Life. Darwin Meets von Neumann*, Vol. 5778 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 158–166.

Foster, S. H.: 1979, Topic initiation: one step or two? some factors involved in mother-child interaction at the prelinguistic stage, *Nottingham Linguistic Circular* **8**(2), 96–107.

Fotion, N. G.: 1975, Indicating Devices?, *Philosphy & Rhetoric* **8**(4), 230–237.

Francis, H.: 1979, What does a child mean? A critique of the functional approach to language acquisition, *Journal of Child Language* **6**, 201–210.

French, L. and Pak, M. K.: 1991, Mothers and Peers as Conversational Partners: Quantity and Quality of Talk. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Seattle, WA, April 18-20.

Garvey, C.: 1983, Text, context, and interaction in language acquisition, *in* R. M. Golinkoff (ed.), *The Transition from Prelinguistic to Linguistic Communication*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp. 191–218.

Goddard, C.: 2006, Ethnopragmatics: a new paradigm, *in* C. Goddard (ed.), *Ethnopragmatics. Understanding Discourse in Cultural Context*, De Gruyter Mouton, Berlin, Boston, chapter 1.

Goffman, E.: 1961, *Encounters: Two Studies in the Sociology of Interaction*, Penguin, Harmondsworth.

Goffman, E.: 1974, *Frame Analysis: An Essay on the Organization of Experience*, Penguin, Harmondsworth.

Golinkoff, R. M.: 1993, When is communication a 'meeting of minds'?, *Journal of Child Language* **20**, 199–209.

Gopnik, A.: 1988, Three types of early word: the emergence of social words, names and cognitive-relational words in the one-word stage and their relation to cognitive development, *First Language* **8**(22), 49–70.

Gosling, S. D., Rentfrow, P. J. and Swann Jr., W. B.: 2003, A Very Brief Measure of the Big Five Personality Domains, *Journal of Research in Personality* **37**, 504–528.

Grice, H. P.: 1957, Meaning, *The Philosophical Review* **66**(3), 377–388.

Grove, W. M., Andreasen, N. C., McDonald-Scott, P., Keller, M. B. and Shapiro, R. W.: 1981, Reliability Studies of Psychiatric Diagnosis: Theory and Practice, *Archives of General Psychiatry* **38**(4), 408–413.

Gullberg, M., de Boot, K. and Volterra, V.: 2008, Gestures and some key issues in the study of language development, *Gesture* **8**(2), 149–179.

Gwet, K.: 2002, Kappa Statistic is not Satisfactory for Assessing the Extent of Agreement Between Raters, *Statistical Methods for Inter-Rater Reliability Assessment* **1**(6), 1–6.

Halliday, M. A. K.: 1975, *Learning How to Mean: Explorations in the Development of Language*, Edward Arnold, London.

Harding, C. G.: 1983, Setting the stage for language acquisition: Communication development in the first year, *in* R. M. Golinkoff (ed.), *The Transition from Prelinguistic to Linguistic Communication*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp. 93–114.

Harnad, S.: 1990, The symbol grounding problem, *Physica D* **42**, 335–346.

Hastie, T., Tibshirani, R. and Friedman, J.: 2013, *The Elements of Statistical Learning*, 2nd edn, Springer.

Hatch, D. J.: 1999, Chomskian Linguistics: God's Truth or Hocus Pocus?, *Ethnographic Studies* **4**, 27–49.

Hayes, A.: 1984, Interaction, engagement, and the origins and growth of communication: Some constructive concerns, *in* L. Feagans, C. Garvey and R. Golinkoff (eds),

The Origins and Growth of Communication, Ablex Publishing Corporation, New Jersey, pp. 136–161.

Hirsh-Pasek, K. and Golinkoff, R. M.: 2012, How babies talk: Six principles of early language development, *in* S. L. Odom, E. P. Pungello and N. Gardner-Neblett (eds), *Infants, Toddlers, and Families in Poverty: Research Implications for Early Child Care*, The Guilford Press, chapter 4, pp. 77–100.

Holmlund, C.: 1995, Development of turntaking as a sensorimotor process in the first 3 months: A sequential analysis, *in* K. E. Nelson and Z. Réger (eds), *Children's Language*, Vol. 8, Lawrence Erlbaum Associates, Inc., New Jersey, pp. 41–64.

Hutchby, I. and Wooffitt, R.: 1999, *Conversation analysis: principles, practices, and applications*, Blackwell.

iCub Forward Kinematics: 2012, `http://eris.liralab.it/wiki/ICubForwardKinematics`, last visited 04th of October.

ITALK project: 2013, `http://www.italkproject.org/`, last visited 9th of September.

Jefferson, G.: 1989, Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation, *in* D. Roger and P. Bull (eds), *Conversation: an interdisciplinary perspective*, Multilingual Matters, Clevedon, chapter 8, pp. 166–196.

Johnson, C. E.: 1982, Children's Questions and the Discovery of Interrogative Syntax: A dissertation, Ann Arbor Mich: University Microfilms International.

Jones, S. and Zimmerman, D. H.: 2003, A child's point and the achievement of intentionality, *Gesture* **3**, 155–185.

Jurafsky, D. and Martin, J. H.: 2000, *Speech and Language Processing*, Prentice Hall, New Jersey.

Kaye, K.: 1977, Toward the origin of dialogue, *in* H. R. Schaffer (ed.), *Studies in Mother-Infant Interaction*, Academic Press, New York, pp. 89–118.

Kaye, K.: 1979, Thickening thin data: The maternal role in developing communication and language, *in* M. Bullowa (ed.), *Before Speech: The Beginning of Interpersonal Communication*, Cambridge University Press, Cambridge, pp. 191–206.

Kaye, K. and Charney, R.: 1980, How mothers maintain dialogue with two-year-olds, *in* D. R. Olsen (ed.), *The Social Foundations of Language & Thought: Essays in Honor of Jerome S. Bruner*, Norton, pp. 211–230.

Kaye, K. and Charney, R.: 1981, Conversational asymmetry between mothers and children, *Journal of Child Language* **8**, 35–50.

Keenan, E. O. and Schieffelin, B. B.: 1976, Topic as a discourse notion: A study of topic in the conversation of children and adults, *in* C. N. Li (ed.), *Subject and Topic*, Academic Press, New York, pp. 337–384.

Kidwell, M. and Zimmerman, D. H.: 2007, Joint attention as action, *Journal of Pragmatics* **39**(3), 592–611.

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M. and Campo, E. D.: 2011, The Representation of Abstract Words: Why Emotion Matters, *Journal of Experimental Psychology* **140**(1), 14–34.

Krippendorff, K.: 1980, *Content Analysis: An Introduction to Its Methodology*, Sage Publications, Newbury Park, California.

Landauer, T. K. and Dumais, S. T.: 1997, A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge, *Psychological Review* **104**(2), 211–240.

Leavens, D. A. and Hopkins, W. D.: 1999, The whole-hand point: The structure and function of pointing from a comparative perspective, *Journal of Comparative Psychology* **113**(4), 417–425.

Leech, G., Rayson, P. and Wilson, A.: 2001, *Word Frequencies in Written and Spoken English: based on the British National Corpus*, Longman.

Levinson, S. C.: 1983, *Pragmatics*, Cambridge University Press.

Levinson, S. C.: 1995, Interactional biases in human thinking, *in* E. N. Goody (ed.), *Social intelligence and interaction*, Cambridge University Press, Cambridge, UK, chapter 11, pp. 221–260.

Levinson, S. C.: 2006, On the Human "Interaction Engine", *in* N. J. Enfield and S. C. Levinson (eds), *Roots of human sociality: culture, cognition and interaction*, Berg Publishers, chapter 1, pp. 39–69.

Lindquist, K. A., Quigley, K. S., Siegel, E. H. and Barrett, L. F.: 2013, The Hundred-Year Emotion War: Are Emotions Natural Kinds or Psychological Constructions? Comment on Lench, Flores, and Bench (2011), *Psychological Bulletin* **139**(1), 255–263.

Liszkowski, U.: 2006, Infant pointing at twelve months: Communicative goals, motives, and social-cognitive abilities, *in* N. Enfield and S. Levinson (eds), *The Roots of Human Sociality: Culture, Cognition, and Interaction*, Berg, Oxford, UK, pp. 153–178.

Liszkowski, U., Carpenter, M. and Tomasello, M.: 2007, Reference and attitude in infant pointing, *Journal of Child Language* **34**, 1–20.

Lock, A.: 1978, *Action, Gesture and Symbol*, Academic Press, London.

Locke, J. L.: 1993, *The Child's Path to Spoken Language*, Harvard University Press, Cambridge, Massachusetts.

Lungarella, M., Metta, G., Pfeiffer, R. and Sandini, G.: 2003, Developmental robotics: a survey, *Connection Science* **15**(4), 151–190.

Lyon, C., Nehaniv, C. L. and Cangelosi, A. (eds): 2007, *Emergence of Communication and Language*, Springer, London.

Lyon, C., Nehaniv, C. L. and Saunders, J.: 2012, Interactive Language Learning by Robots: The Transition from Babbling to Word Forms, *PLoS ONE* **7**(6).

Markman, E. M.: 1990, Constraints Children Place on Word Meanings, *Cognitive Science* **14**(1), 57–77.

Merritt, M.: 1976, On questions following questions (in service encounters), *Language in Society* **5**(3).

Meumann, E.: 1902, Die Entstehung der ersten Wortbedeutungen beim Kinde, *Philosophische Studien*, p. 20.

Murray, L. and Trevarthen, C.: 1986, The infant's role in mother-infant communications, *Journal of Child Language* **13**, 15–29.

Nehaniv, C. L., Dautenhahn, K. and Loomes, M. J.: 1999, Constructive Biology and Approaches to Temporal Grounding in Post-Reactive Robotics, *in* G. T. McKee and

P. Schenker (eds), *Sensor Fusion and Decentralized Control in Robotics Systems II*, Vol. 3839 of *Proc. of SPIE*, pp. 156–167.

Nicholas, J. G., Geers, A. E. and Rollins, P. R.: 1999, Inter-rater reliability as a reflection of ambiguity in the communication of deaf and normally-hearing children, *Journal of Communication Disorders* **32**(2), 121–134.

Ninio, A., Snow, C. E., Pan, B. A. and Rollins, P. R.: 1994, Classifying communicative acts in children's interactions, *Journal of Communication Disorders* **27**(2), 157–187.

OED online: 2013, "yes, adv. and n.2", `http://www.oed.com/view/Entry/231637` last visited 17th of May 2013.

Owens, R. E.: 2012, *Language Development: An Introduction*, 8 edn, Pearson Education.

Pattacini, U.: 2010, *Modular Cartesian Controllers for Humanoid Robots: Design and Implementation on the iCub*, PhD thesis, RBCS, Istituto Italiano di Tecnologia, Genoa, Italy.

Pattacini, U., Nori, F., Natale, L., Metta, G. and Sandini, G.: 2010, An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots, *Int. Conf. on Intelligent Robots and Systems (IROS)*, IEEE/RSJ, IEEE, pp. 1668–1674.

Pea, R.: 1980, The Development Of Negation In Early Child Language, *in* D. Olson (ed.), *The Social Foundations of Language & Thought: Essays in Honor of Jerome S. Bruner*, W.W. Norton, pp. 156–186.

Piaget, J.: 1954, *The Language and Thought of the Child*, Norton, New York.

Piaget, J. and Cook, M. T.: 1952, *The origins of intelligence in children*, W.W. Norton & Company, New York.

Piatelli-Palmarini, M. (ed.): 1980, *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*, Harvard University Press, Cambridge, Massachusetts.

Poulin-Dubois, D. and Graham, S. A.: 2007, Cognitive Processes in Early Word Learning, *in* E. Hoff and M. Shatz (eds), *Blackwell Handbook of Language Development*, Blackwell Publishing, chapter 10, pp. 191–211.

Protection from Harassment Act 1997: 1997, `http://www.legislation.gov.uk/ukpga/1997/40`, last visited 16th of March 2013.

Quinlan, J. R.: 1992, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA.

Rawls, A.: 2003, Harold Garfinkel, *in* G. Ritzer (ed.), *The Blackwell Companion to Major Contemporary Social Teorists*, Wiley-Blackwell.

Rietveld, T. and van Hout, R.: 1993, *Statistical Techniques For the Study of Language and Language Behaviour*, Mouton de Gruyter, Berlin.

RobotCub project: 2013, `http://www.robotcub.org/`, last visited September.

Roy, B. C., Frank, M. C. and Roy, D.: 2009, Exploring word learning in a high-density longitudinal corpus, *Proceeding of the Thirty-First Annual Conference of the Cognitive Science Society*.

Roy, D.: 2005, Semiotic schemas: A framework for grounding language in action and perception, *Artificial Intelligence* **167**(1–2), 170–205.

Rüsch, J., Lopes, M., Bernardino, A., Hörnstein, J. and Santos-Victor, J.: 2008, Multimodal saliency-based bottom-up attention - a framework for the humanoid robot icub,

2008 IEEE International Conference on Robotics and Automation, ICRA 2008, May 19-23, 2008, Pasadena, California, USA, IEEE, pp. 962–967.

Ryan, J.: 1974, Early language development: Towards a communicational analysis, *in* M. P. M. Richards (ed.), *The integration of a child into a social world*, Cambridge University Press.

Sacks, H., Schegloff, E. A. and Jefferson, G.: 1974, A Simplest Systematics for the Organization of Turn-Taking for Conversation, *Language* **50**(4), 696–735.

Saunders, J., Lehmann, H., Förster, F. and Nehaniv, C. L.: 2012, Robot Acquisition of Lexical Meaning - Moving Towards the Two-word Stage, *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pp. 1–7.

Saunders, J., Lehmann, H., Sato, Y. and Nehaniv, C. L.: 2011, Towards Using Prosody to Scaffold Lexical Meaning in Robots, *Proceedings of ICDL-EpiRob 2011*.

Saunders, J., Lyon, C., Förster, F., Nehaniv, C. L. and Dautenhahn, K.: 2009, A Constructivist Approach to Robot Language Learning via Simulated Babbling and Holophrase Extraction, *Proc. 2nd Intl. IEEE Symposium on Artificial Life*, pp. 13–20.

Saunders, J., Nehaniv, C. L. and Dautenhahn, K.: 2006, Teaching robots by moulding behaviour and scaffolding the environment, *in* M. A. Goodrich, A. C. Schultz and D. J. Bruemmer (eds), *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot interaction*, ACM, pp. 118–125.

Saunders, J., Nehaniv, C. L., Dautenhahn, K. and Alissandrakis, A.: 2007, Self-imitation and environmental scaffolding for robot teaching, *International Journal of Advanced Robotic Systems* **4**(1), 109–124.

Saunders, J., Nehaniv, C. L. and Lyon, C.: 2010, Robot Learning of Lexical Semantics from Sensorimotor Interaction and the Unrestricted Speech of Human Tutors, *Proc. of the Second International Symposium on New Frontiers in Human-Robot Interaction at AISB convention 2010, Leicester, UK*, pp. 95–102.

Saunders, J., Nehaniv, C. L. and Lyon, C.: 2011, The acquisition of word semantics by a humanoid robot via interaction with a human tutor, *in* K. Dautenhahn and J. Saunders (eds), *New Frontiers in Human-Robot Interaction*, John Benjamins Publishing Company, pp. 211–234.

Schaffer, H. R.: 1979, Acquiring the concept of dialogue, *in* M. H. Bornstein and W. Kessen (eds), *Psychological Development from Infancy: Image to Intention*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp. 279–305.

Schaffer, H. R.: 1984, *The Child's Entry into a Social World*, Academic Press, London.

Schegloff, E. A.: 1968, Sequencing in conversational openings, *American Anthropologist* **70**, 1075–95.

Schegloff, E. A.: 1991, Conversation analysis and socially shared cognition, *in* L. Reshnick, J. Levine and S. Teasley (eds), *Perspectives on Socially Shared Cognition*, American Psychological Association, Washington DC, pp. 150–171.

Schegloff, E. A.: 1992, On talk and its institutional occasions, *in* P. Drew and J. Heritage (eds), *Talk at Work: Interaction in Institutional Settings*, Cambridge University Press, Cambridge, pp. 101–134.

Schegloff, E. A. and Sacks, H.: 1973, Opening up closings, *Semiotica* **7**, 289–327.

Schieffelin, B. B. and Ochs, E.: 1983, A cultural perspective on the transition from prelinguistic to linguistic communication, *in* R. M. Golinkoff (ed.), *The Transition from Prelinguistic to Linguistic Communication*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp. 115–132.

Searle, J. R.: 1969, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press.

Searle, J. R., Parret, H. and Verschueren, J. (eds): 1992, *(On) Searle on Conversation*, John Benjamins Publishing Company.

Sharrock, W. and Anderson, B.: 1987, Work Flow in a Paediatric Clinic, *in* G. Button and J. R. E. Lee (eds), *Talk and Social Organization*, Multilingual Matters, Clevedon, pp. 244–260.

Siskind, J.: 2001, Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic, *Journal of Artificial Intelligence Research* **15**, 31–90.

Snow, C. E.: 1977, The development of conversation between mothers and babies, *Journal of Child Language* **4**, 1–22.

Snow, C. E.: 1978, The conversational context of language acquisition, *in* R. Campbell and P. Smith (eds), *Recent Advances in the Psychology of Language: Language Development and Mother-Child Interaction*, Plenum Press, New York, pp. 253–269.

Snow, C. E., Perlmann, R. and Nathan, D.: 1987, Why Routines are different: Towards a multiple factors model of the relation between input and language acquisition, *in* K. E. Nelson and A. van Kleeck (eds), *Children's Language*, Vol. 6, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp. 65–97.

Spitz, R. A.: 1957, *No and Yes: On the Genesis of Human Communication*, International Universities Press.

Steels, L.: 1998, The origins of syntax in visually grounded robotic agents, *Artificial Intelligence* **103**(1-2), 133–156.

Steels, L.: 2003, Evolving grounded communication for robots, *Trends in Cognitive Science* **7**(7), 308–312.

Steels, L.: 2005, The emergence and evolution of linguistic structure: from lexical to grammatical communication systems, *Connection Science* **17**(3-4), 213–230.

Steels, L. and Baillie, J.: 2003, Shared grounding of event descriptions by autonomous robots, *Robotics and Autonomous Systems* **43**, 163–173.

Steels, L. and Kaplan, F.: 2002, AIBO's first words: The social learning of language and meaning, *Evolution of Communication* **4**(1), 3–32.

Stella-Prorok, E. M.: 1983, Mother-child language in the natural environment, *Children's Language Vol. 4*, Gardner Press, New York, pp. 187–230.

Stern, D. N.: 1974, Mother and infant at play: The dyadic interaction involving facial, vocal and gaze behaviours, *in* M. Lewis and L. A. Rosenblum (eds), *The Effect of the Infant on its Caregiver*, John Wiley and Sons, New York, pp. 187–213.

Symons, D. and Moran, G.: 1987, The behavioral dynamics of mutual responsiveness in early face-to-face mother-infant interaction, *Child Development* **58**(6), 1488–1495.

Tanz, C.: 1987, Introduction, *in* S. U. Phillips, S. Steele and C. Tanz (eds), *Language, Gender, and Sex in Comparative Perspective*, Cambridge University Press, Cambridge, pp. 163–177.

te Molder, H. and Potter, J. (eds): 2005, *Conversation and Cognition*, Cambridge University Press, Cambridge.

ten Have, P.: 2007, *Doing Conversation Analysis*, 2nd edn, Sage Publications.

Tideman, T. N.: 1987, Independence of Clones as a Criterion for Voting Rules, *Social Coice and Welfare* **4**(3), 185–206.

TiMBL: 2012, `http://ilk.uvt.nl/mblp/`, last visited 05th of October.

Tomasello, M.: 2003, *Constructing a Language*, Harvard University Press.

Tomasello, M., Carpenter, M. and Liszkowski, U.: 2007, A New Look at Infant Pointing, *Child Development* **78**(3), 705–722.

Trevarthen, C.: 1979, Communication and co-operation in early infancy, *in* M. Bullowa (ed.), *Before Speech: The Beginning of Interpersonal Communication*, Cambridge University Press, Cambridge, pp. 321–347.

Trevarthen, C. and Aitken, K. J.: 2001, Infant Intersubjectivity: Research, Theory, and Clinical Applications, *Journal of Child Psychology and Psychiatry* **42**(1), 3–48.

Underhill, J. W.: 2011, *Creating Worldviews*, Edinburgh University Press.

Varela, F. J., Thompson, E. and Rosch, E.: 1991, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, Cambridge, Massachusetts.

Vygotsky, L.: 1986, *Thought and Language*, MIT Press, Cambridge, Massachusetts.

Weisstein, E. W.: 2013, Zipf Distribution, From MathWorld – A Wolfram Web Resource `http://mathworld.wolfram.com/ZipfDistribution.html`, last visited 15th of May 2013.

Werner, H. and Kaplan, B.: 1963, *Symbol Formation*, Wiley, New York.

Wierzbicka, A.: 1994, "Cultural Scripts": A Semantic Approach to Cultural Analysis and Cross-Cultural Communication, *Pragmatics and Language Learning. Monograph Series* **5**, 1–24.

Wittgenstein, L.: 1958, *Philosophical Investigations*, 3 edn, Blackwell, Oxford.

Wittgenstein, L.: 1984, Philosophische Untersuchungen, *Werkausgabe - Band 1. Tractatus logico-philosophicus*, Suhrkamp.

Wyllys, R. E.: 1981, Empirical and theoretical bases of Zipf's law, *Library Trends* **30**(1), 53–64.

Yoder, P., Davis, B., Bishop, K. and Munson, L.: 1994, Effect of Adult Continuing Wh-Questions on Conversational Participation in Children With Developmental Disabilities, *Journal of Speech and Hearing Research* **37**(1), 193–204.

Zukow, P. G., Reilly, J. and Greenfield, P. M.: 1982, Making the absent present: Facilitating the transition from sensorimotor to linguistic communication, *in* K. E. Nelson (ed.), *Children's Language Vol. 3*, Lawrence Erlbaum Associates Inc., Hillsdale, NJ, pp. 1–90.