THE ROLE OF SEMANTIC COMPLEXITY FOR THE ACQUISITION OF ADJECTIVES

Inauguraldissertation

zur Erlangung des Grades eines Doktors der Philosophie

im Fachbereich Neuere Philologien

der Johann Wolfgang Goethe-Universität

zu Frankfurt am Main

vorgelegt von

Merle Weicker

aus

Heppenheim (Bergstraße)

Einreichungsjahr: 2018

Erscheinungsjahr: 2019

1. Gutachter: Prof. Dr. Petra Schulz
2. Gutachter: Prof. Dr. Thomas Ede Zimmermann
Tag der mündlichen Prüfung: 18. April 2019

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. Petra Schulz for sparking my interest in language acquisition, for her ideas, support, and feedback before and during my time as a PhD student, and for encouraging me to present my research to others. I am also very grateful to Prof. Dr. Ede Zimmermann for his openness to bring formal semantic theory and language acquisition research together. This thesis benefitted greatly from his valuable comments on the theoretical aspects of nominal modification.

I also thank the current and former members of the graduate school *GRK 2016/1 "Nominal Modification"* and the language acquisition team at Goethe-University Frankfurt for numerous discussions about empirical and theoretical approaches on adjectives. I wish to thank Cécile Meier and Manfred Sailer for their interesting classes on the semantics of adjectives, Sascha Alexeyenko for the discussion group as well as Rabea Lemmer and Corinna Trabandt for their comments on this thesis. Moreover, I am much obliged to my colleagues for their organizational and moral support. Sabrina Geyer, Eugenia Greco, Alina Lausecker, Rabea Lemmer, Corinna Trabandt, and Caro Reinert, thanks for your advices!

These great working conditions gave me the opportunity to meet inspiring researchers from other universities. In this context I thank Ana Pérez-Leroux and her team, Tom Roeper, and Kristen Syrett for their helpful feedback at different stages of my dissertation project. A Minerva Short Term Research Grant afforded me to visit Galit Sassoon, Natalia Meir, Aya Meltzer-Asscher, and Sharon Armon-Lotem at Tel Aviv University and Bar-Ilan University. I am grateful to all of them for sharing their knowledge with me and for their hospitality.

I would also like to thank the child and adult participants for taking part in the experiment and the kindergartens for their cooperation. Many thanks go to Valeria Bunkov, Marina Kukina, Alex Lowles, and Silvana Pfuhz for helping me with the data collection. I am also really thankful to Alex Lowles for drawing the stimuli of the experiment.

Most of all I thank my family, my friends, and Leon, who is both, for their intuition when emotional support was needed, for taking the load off me, and for reminding me that there is more to life than adjectives.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The following utterances come from a monolingual German-speaking child who is not even 4[1]:

(1.1) a. *CHI (2;00 years)[2]: **braune** Enten
'brown ducks'

b. *CHI (2;01 years): **vollen** Teller Nudeln
'full plate of pasta'

c. *CHI (2;01 years): ein **kleiner** Tunnel
'a small tunnel'

d. *CHI (2;10 years): ja, was machen unseren **kranken** Tiere?
'what about our sick animals?'

e. *CHI (2;10 years): ein ziemlich **hässlicher** Autofahrer
'a fairly ugly car driver'

f. *CHI (3;04 years): das wird das **frühere** Florenz
'this is going to be the Florence of former times'

The examples in (1.1) illustrate that at age 2, children produce various adjectives to specify the noun phrase referent. Do some adjectives occur earlier than others in modifier position as indicated by the age at which the utterances in (1.1) are produced? Do adjectives such as in example (1.1a) to (1.1f) differ? If yes, are these differences mirrored in the acquisition of adjectives? I argue that the adjectives in (1.1) differ with respect to compositional and lexical semantic properties. In particular, the differences concern entailment properties, i.e., intersectivity and subsectivity, and gradability properties, i.e., non-gradable versus gradable, absolute versus relative gradable, dimensional versus evaluative.

The adjective-noun phrases in examples (1.1a) to (1.1e) license the inference that both the property denoted by the adjective and the property denoted by the noun is true for

[1] The data are taken from the "Leo" corpus (Behrens, 2006), which is available at the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000).

[2] In the remainder of the thesis the children's age is given in the format *years;months*. The asterix in front of a speaker name is part of the CHILDES notation and does not mark ungrammaticality.

the noun phrase referent. For instance, *a brown duck* is an entity that is brown and a duck. This inference is possible because the adjectives in these examples denote sets of individuals like nouns such as *duck* do. Hence, they are INTERSECTIVE, i.e., they can be combined via the intersection of the set denoted by the adjective and the set denoted by the noun. Intersective adjectives are always SUBSECTIVE. This means the set of entities denoted by the adjective-noun phrase is a subset of the set of entities denoted by the noun. For instance, the set of brown ducks is a subset of the set of ducks. These inferences are not possible for the adjective-noun phrase in example (1.1f) because *former* does not denote a set of individuals. Moreover, the sentence implies that the entity denoted by the adjective-noun phrase is not a subset of the set of the cities called 'Florence'[3] at the time the sentence is uttered. These differences relate to the compositional mechanisms involved in the combination of adjective and noun. In addition, the adjectives in examples (1.1a) to (1.1f) cannot be modified by the same degree adverbs, which points to differences in the adjectives' gradability properties. *Brown* in (1.1a) and *former* in (1.1f) cannot be modified by degree adverbs such as *very* because they are not gradable. In contrast, this modification is possible for *full, small, sick*, and *ugly* in (1.1b) to (1.1e). These four adjectives are gradable. Unlike non-gradable adjectives such as *brown* and *former*, gradable adjectives must be interpreted with respect to a standard of comparison. For RELATIVE GRADABLE adjectives (e.g., *small, ugly*) the standard of comparison is dependent on a comparison class, for ABSOLUTE GRADABLE adjectives (e.g., *full, sick*) the standard is independent of a comparison class. This difference is related to the adjectives scale structure, which is encoded in the lexicon (e.g., Kennedy & McNally, 2005). Absolute and relative gradable adjectives exist for DIMENSIONAL and EVALUATIVE adjectives. Both evaluative adjectives (e.g., *sick, ugly*) and dimensional adjectives (e.g., *full, small*) encode the adjective's scale structure in the lexicon. However, evaluative adjectives encode additional properties, for instance that they can receive a subjective interpretation. I claim that the different compositional and lexical semantic properties of adjectives result in differences in the adjectives' semantic complexity. Given the different semantic properties of adjectives, this thesis addresses the following research question:

(Q) How do monolingual German-speaking children acquire the semantics of adjectives?

My findings from a longitudinal speech data analysis and a comprehension experiment suggest a relation between the adjectives' semantic complexity and their acquisition. The findings indicate that children acquire the semantics of adjectives in an order determined by semantic complexity: more complex adjectives are not acquired before less complex adjectives and adjectives of the same complexity are acquired simultaneously. These findings contribute to research on both language acquisition and formal semantic theory. This thesis makes a contribution to semantic theory because it attempts the integration of different properties of adjectives. In addition, I suggest a definition of semantic complexity

[3] Note that in example (1.1f) it is possible that *Florence* refers to different historical stages of the city of Florence rather than to the set of cities called 'Florence'.

– a notion that has been addressed only rarely (Matthewson, 2014). The empirical data I obtained by psycholinguistic experiments with adults and children provide support for the assumed semantics of adjectives (Castroviejo, McNally, & Sassoon, 2018; Van Geenhoven, 2006). What is more, this thesis contributes to the common question in the field of lexical acquisition of which factors play a role in the acquisition order of lexical categories (Blackwell, 2005).

In acquisition research, adjectives have received less attention than nouns and verbs. Research on adjective semantics and interpretation is especially scarce. This may be due to methodological and theoretical reasons. As for methodological reasons, it is in general difficult to find suitable research designs for addressing the question of how children acquire the meaning of words (see Rothweiler, 2007; Van Geenhoven, 2006). In addition, adjectives are produced less frequently than nouns and verbs (Kauschke, 2000). As for theoretical reasons, it can become difficult to isolate adjectives from other word classes because they often share properties with for instance nouns and verbs (Cabredo Hofherr, 2010; Dixon, 2004). Second, adjectives can occur in different syntactic structures. In the examples in (1.1) the adjectives are in attributive position, i.e., they appear prenominally. However, most adjectives can also appear in predicative position as in *Die Ente ist braun* ('The duck is brown'). It is still under discussion whether there is a syntactic and semantic relationship between theses structures and how it may look like. Most importantly, linguistic theory has provided diverse classification systems according to which adjectives should be categorized semantically: adjectives have been classified according to notional properties that express concepts such as COLOR, DIMENSION, AGE etc. (Dixon, 1982, 2004), according to entailment properties such as intersective or non-intersective (e.g., Kamp & Partee, 1995), and according to gradability properties such as gradable and non-gradable (e.g., Bierwisch, 1989). Gradable adjectives have been further categorized into closed scale versus open scale adjectives, absolute versus relative gradable adjectives, dimensional versus evaluative adjectives, and uni- versus multidimensional adjectives (e.g., Bierwisch, 1989; Kennedy & McNally, 2005; Sassoon, 2013).

As a consequence of the methodological and theoretical challenges, previous studies have focused on selected properties of adjectives. Studies focusing on lexical properties of adjectives investigated whether children are able to distinguish between novel words that refer either to properties or to kinds of objects by using morphosyntactic cues, e.g., *this is a blikish one* versus *this is a blik* (e.g., Waxman & Booth, 2001). Other studies focused on the acquisition of specific adjectives, mostly color and dimension words. With regard to dimension words the acquisition of positive and negative pole adjectives (e.g., *big-small*) was investigated (e.g., Siegel, 1977), the comprehension of general (e.g., *big*) and specific (e.g., *thick*) dimensions words (e.g., Eilers, Oller, & Ellington, 1974), the interpretation of relational adjectives like *big* (e.g., Smith, Cooney, & McCord, 1986; Tribushinina, 2013), and the distinction between relative (e.g., *big*) and absolute (e.g., *clean*) gradable adjectives (e.g., Foppolo & Panzeri, 2013; Syrett, Kennedy, & Lidz, 2010). Studies focusing on the compositional semantics of adjectives examined children's abilities to integrate adjective

and noun meaning and to use the adjective to distinguish between referents (Fernald, Thorpe, & Marchmann, 2010). Other studies investigated the role of the noun for the interpretation of relative gradable adjectives (e.g., *big mouse* versus *big elephant*) (Barner & Snedeker, 2008). Another line of research is concerned with the factors influencing the order of acquisition of adjectives. The following factors have been proposed: properties of the input (Sandhofer & Smith, 2007; Tribushinina et al., 2014), the adjective's dependency on the noun meaning (Blackwell, 2005), and the adjectives' complexity defined as the concreteness or abstractness of the underlying concept (Tribushinina et al., 2014).

In summary, semantic properties of adjectives have been mostly investigated independently of each other in both theory and acquisition research. Therefore, one objective of this thesis is to provide a classification system for adjectives that combines different semantic properties. This system comprises six classes, which constitute a *Semantic Complexity Hierarchy*. The first division of adjectives concerns the mechanism by which the adjective modifier and the head noun combine and classifies adjectives as subsective/intersective (e.g., *brown, full, small, sick, ugly*) and non-subsective/non-intersective (e.g., *former*). Intersective adjectives are further specified according to their gradability properties. This results in the distinction of gradable (e.g., *full, small, sick, ugly*)[4] and non-gradable adjectives (e.g., *brown*). Evaluative (e.g., *sick, ugly*) and dimensional (e.g., *full, small*) adjectives are subclasses of gradable adjectives. Both classes encode the adjectives' scale structure in the lexicon resulting in the distinction of relative (e.g., *small, ugly*) and absolute (e.g., *full, sick*) gradable adjectives. Evaluative adjectives encode additional properties in the lexicon, for instance that they can receive a subjective interpretation.

To the best of my knowledge, I propose the first operationalization of semantic complexity with respect to adjectives, which comprises the compositional and lexical semantic properties that give rise to differences between the adjectives in (1.1). This definition of semantic complexity results in the following hierarchy with complexity increasing from intersective/subsective/non-gradable adjectives to non-intersective/non-subsective adjectives:

(1.2) intersective/subsective/non-gradable (e.g., *brown*)
$<$ intersective/subsective/absolute gradable dimensional (e.g., *full*)
$=$ intersective/subsective/relative gradable dimensional (e.g., *small*)
$<$ intersective/subsective/absolute gradable evaluative (e.g., *sick*)
$=$ intersective/subsective/relative gradable evaluative (e.g., *ugly*)
$<$ non-intersective/non-subsective (e.g., *former*)

In contrast to previous studies on the acquisition of adjectives, I provide a formal definition of complexity and propose that semantic complexity as stated in the *Semantic Complexity Hierarchy* determines the order of acquisition of adjectives. Consequently, I propose that the input is not the major factor determining the acquisition order. I test

[4] Note that gradable adjectives are sometimes analyzed as non-intersective. I follow the approaches that analyze them as intersective for reasons discussed in Chapter 2.

these two proposals in two studies investigating the production of adjectives in spontaneous speech and their comprehension in a controlled experimental setting in monolingual German-speaking children. The studies span the age range between two and five years. I chose this age range because lexical acquisition starts early, but the specification of all meaning components of a word is a lengthy process (Müller, Schulz, & Tracy, 2018). The empirical studies of this thesis contribute to a more comprehensive picture regarding the acquisition of adjectives without the restriction to one particular property of adjectives. Moreover, the findings of the empirical studies allow to postulate a developmental path for the acquisition of semantic properties of adjectives. In addition, this acquisition path considers findings from previous studies concerning the acquisition of notional properties of adjectives, children's cognitive abilities, and possible triggers in the input.

This thesis is structured as follows. Chapter 2 gives an overview over the classification of adjectives. I start with the syntactic structures in which adjectives can occur. Regarding the semantic classifications, I present approaches that classify adjectives either according to notional properties, according to entailment properties, or according to gradability properties. On the basis of the integration of entailment- and gradability properties, I propose six different adjective classes that differ in semantic complexity and derive the first acquisition hypothesis that semantic complexity determines the acquisition order of adjectives.

Chapter 3 summarizes previous acquisition research on existing and novel adjectives occurring in different syntactic structures and on gradable adjectives. Studies on the acquisition of notional adjective classes address the question of the acquisition order of adjectives. These studies propose that properties of the input such as adjective frequency influence the acquisition order of adjectives. However, previous findings regarding the role of the input for the acquisition of adjectives are mixed. Due to these inconclusive results I derive the second acquisition hypothesis that the acquisition order of adjectives is not completely determined by the input.

In Chapter 4, I report the spontaneous speech study. This case study is a longitudinal investigation of the production of adjectives from 2;00 to 2;11 years based on transcripts from a dense data corpus. To test the hypothesis that semantic complexity determines the order of acquisition, I consider all classes of the *Semantic Complexity Hierarchy*. I analyze the parental input as an alternative determining factor. The results provide evidence that the mean age of acquisition for the respective adjective classes follows the order predicted by semantic complexity. The same order was observed for the age at which the number of types for each class increased most. A preliminary analysis of the input indicates that the frequency of parental adjective use is related to the order of acquisition, but it is unlikely that frequency determines the order completely. Besides the semantic properties of the adjectives produced by the child, I address the adjectives' notional properties and the syntactic structures in which the adjectives occur. I present preliminary ideas concerning the relation between semantic, conceptual, and syntactic properties.

In Chapter 5, I present the comprehension experiment. This experiment focuses on two specific adjective classes. I examine children's and adults' interpretation of relative (*big,*

small) and absolute (*clean, dirty*) gradable dimensional adjectives with a forced picture-choice task. I investigate whether children distinguish between relative and absolute gradable adjectives when calculating the standard of comparison and which gradable adjectives they interpret relative to a comparison class. In contrast to most previous studies, the comparison class is encoded linguistically by the noun, and hence was made explicit. A target-like interpretation for absolute gradable adjectives is a standard at the endpoint of the scale. Changes in the comparison class, i.e., the noun, should not affect the standard of comparison. The standard for relative gradable adjectives should be around the center of the scale. Changes in the comparison class are expected to affect the standard of comparison. Because relative and absolute gradable adjectives are of the same complexity, I predict that they are acquired, i.e., interpreted target-like, at the same age. The results suggest that 4- and 5-year-old children do not differ from adults in their interpretation of both absolute and relative gradable adjectives. The findings also indicate developmental differences in the comprehension of gradable adjectives because the group of 3-year-old children do not show complete mastery. The younger children are able to locate the standard at the adult-like position on the scale, but the comparison class has the same non-adult-like relevance for relative and absolute gradable adjectives.

In Chapter 6, I discuss the findings from the production and the comprehension study. I summarize the findings and relate them to the question of which factors determine the order of acquisition of adjectives and conclude that semantic complexity plays an influential role. I integrate the empirical findings of this thesis with the findings from semantic theory and previous acquisition studies and propose a developmental path for the acquisition of adjectives. I argue that children enter the acquisition process with linguistic predispositions, e.g., the *Semantic Complexity Hierarchy*, and cognitive predispositions to categorize the environment. I suggest that initially, children apply the least complex interpretation available in the *Semantic Complexity Hierarchy* to all adjectives. This means at this acquisition stage, all adjectives are interpreted as properties of individuals that are not gradable. To access other levels of the *Semantic Complexity Hierarchy* and to establish more complex adjective classes, positive evidence from the input and notional properties of adjectives can operate as triggers. I also discuss implications for different classification systems of adjectives, for the acquisition of adjectives across different languages, and for the acquisition of other modification structures.

In Chapter 7, I summarize the main findings and suggest ideas for future research.

Chapter 2

Syntax and semantics of adjectives

This chapter provides the theoretical background for the acquisition studies presented in Chapters 4 and 5. Section 2.1 serves as an overview over the structures in which adjectives can occur. Section 2.2 concerns the semantics of adjectives. It takes those aspects into account that are relevant for the acquisition data. Other aspects like (non-)restrictivity or the stage-level/individual-level distinction are also part of the semantics of adjectives, but are not included in this chapter because these aspects are not addressed in the empirical part of this thesis.

The theoretical approaches to the syntax and semantics of adjectives presented in this chapter can provide the basis for predictions regarding the acquisition of adjectives. On the other hand, there are still unresolved questions regarding the formal analysis of adjectives or at least aspects that are controversial. Here, theory may benefit from acquisition data in answering these questions.

2.1 Functions of adjectives and syntactic reflexes

Across languages adjectives typically express two functions (e.g., Dixon, 2004): adjectives either describe a property of an entity as in (2.1) or they specify the set of possible referents as in (2.2).

(2.1) Dumbo ist grau-ø.
Dumbo be.3SG gray-ø
'Dumbo is gray.'

(2.2) Der blau-e Elefant-ø heißt
the.NOM.SG.MASC blue-NOM.SG.MASC elephant-NOM.SG.MASC is called
Heffalump.
Heffalump
'The blue elephant is called Heffalump.'

In (2.1), the adjective is the complement of a copula. This use is called 'predicative'.[1] In (2.2), the adjective is part of the DP[2] as a prenominal modifier of the noun. This use is called 'attributive'. Note that sometimes the term 'attributive' is also used for adjectives that cannot be paraphrased with a copular construction as in example (2.4). I use the term 'attributive' exclusively and for all adjectives in prenominal position. In a sentence like *Heffalump ist ein blauer Elefant* ('Heffalump is a blue elephant') the adjective *blau* is part of the DP, hence attributive, although the entire DP is used predicatively. In German, predicative adjectives are always uninflected. Adjectives that are part of the DP agree with the noun in gender, number, and case. If the referent has previously been introduced to the context, it is also possible to elide the noun as in (2.3).

(2.3) Dumbo und Heffalump sind Elefanten.

Der	blau-e	heißt	Heffalump.
the.NOM.SG.MASC	blau-NOM.SG.MASC	is called	Heffalump

'Dumbo and Heffalump are elephants. The blue one is called Heffalump.'

In other languages such as Italian adjectives appear mostly postnominally or rather can occur both in pre- and postnominal position inside the DP. Theories disagree on whether one position is derived from the other position or whether they have different underlying syntactic structures. Approaches that argue for different syntactic structures motivate the variation in adjective positions by claiming that each position is related to a different interpretation (e.g., Cinque, 2010). Postnominal adjectives assign a property to the noun referent, whereas prenominal adjectives modify the reference of the noun. German only allows adjectives in prenominal position inside the DP, but the interpretations related to the pre- and postnominal position in other languages mirror the functions of German predicative and attributive uses. Therefore, German adjective structures raise a similar question regarding the relation of the two adjective positions. The observation that most adjectives can occur in attributive and predicative position (or in pre- and postnominal position inside the DP) is taken as an argument for assuming a derivational relation between the two adjective positions. Early approaches assume that attributive adjectives are derived from predicative adjectives via a (reduced) relative clause (see Demonte, 2011, for an overview). This means prenominal adjectives originate in a postnominal position as part of a reduced relative clause. By predicate-fronting they cross the subject DP and arrive in a prenominal position. Similarity between attributive modifiers and relative clauses speaks in favor of this analysis. Moreover, it has been argued that predicative adjectives are morphologically

[1] I do not include adjectives occuring in secondary predication as in *Sie malte das Haus blau* ('She painted the house blue') or *Sie verließ wütend das Haus* ('She left the house angry') (examples from Demonte, 2011). I restrict myself to adjectives that constitute the main predicate of the clause with *sein* ('be') as the copula.

[2] I assume that the nominal projection is dominated by a functional projection (DP). An overview of the arguments for a D-head can be found in Alexiadou, Haegeman, and Stavrou (2007).

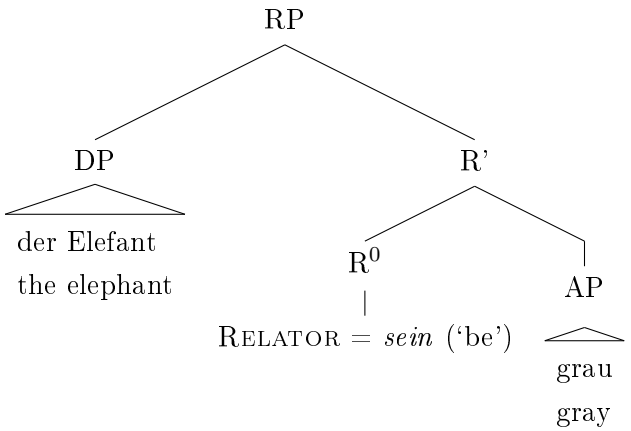
simpler than attributive ones because they are not inflected, and therefore should be the underlying form. However, there are adjectives that cannot be used predicatively such as *ehemalig* ('former').

(2.4) *Der Präsident ist ehemalig.
'The president is former.'

In contrast to the assumption that attributive adjectives are derived from predicative ones, Bolinger (1967) and Montague (1970) assume that predicative adjectives are derived from attributive ones. Syntactically, this means that the adjective is base-generated in a prenominal position. To derive the postnominal position of the adjective, the noun must undergo leftward movement.

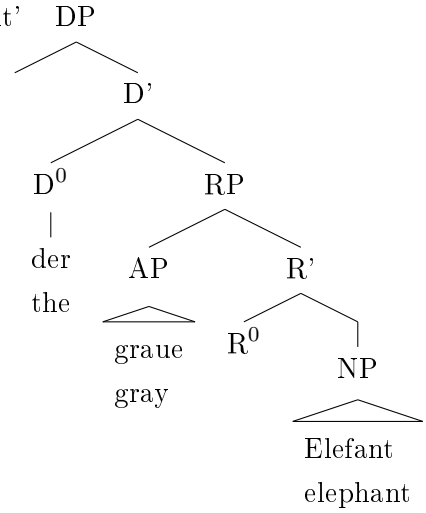
The same configuration for attributive and predicative structures is proposed by den Dikken (2006). He assumes that all subject-predicate relationships are mediated by a RELATOR. The RELATOR is an abstract functional head, but of no designated category. According to den Dikken (2006), it is a placeholder for any functional head. An important assumption of his analysis is that predication relationships are non-directional. Therefore, his analysis can account for both predicative and attributive structures. In predicative structures as in (2.5), the subject is in the specifier of the RELATOR-phrase (RP) and the adjective is in the complement of the RELATOR-head (R^0). The RELATOR-head is occupied by the coupula.

(2.5) Der Elefant ist grau.
'The elephant is gray.'



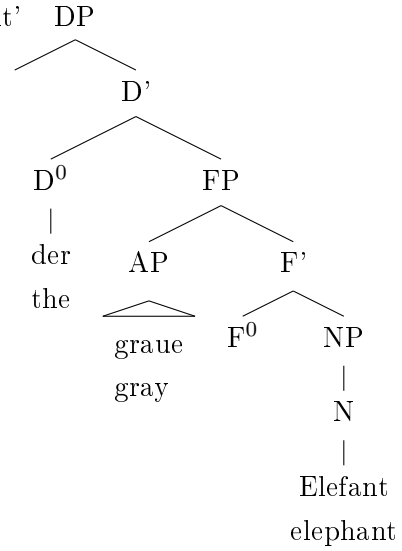
In attributive structures as in (2.6), the predicate is in the specifier of RP and the subject is the complement of the RELATOR-head. In this configuration the RELATOR-head is empty. Because the adjective is in the extended projection of N and the RELATOR can be of any category, this configuration can account for the agreement between adjective and noun.

(2.6) der graue Elefant
'the gray elephant'



Besides the fact that some adjectives cannot occur in predicative position, differences in meaning between attributive and predicative position are another argument against a unified syntactic analysis of the two uses. For instance, the sentence *Mein Freund ist alt* ('My friend is old') means that my friend is of old age, but the sentence *Mein alter Freund* ('My old friend') has the additional meaning that we know each other for a long time. According to den Dikken (2006), this problem can be solved by different lexicalizations of the RELATOR-head. However, the interpretational differences can also be observed with respect to pre- and postnominal occurrences inside the DP (see Cinque, 2010, for an extensive overview). Based on this observation, different underlying structures have been proposed for pre- and postnominal adjectives. According to these approaches, postnominal adjectives are related to a reduced relative clause or predicative structure, whereas prenominal adjectives are related to functional projections intervening between DP and NP. Depending on their meaning, prenominal adjectives reach their position either by predicate-fronting, i.e., they are derived from a reduced relative clause, or they are base-generated in a corresponding functional projection. The resulting structure for a German DP containing a prenominal adjective looks as follows:

(2.7) der graue Elefant
'the gray elephant'



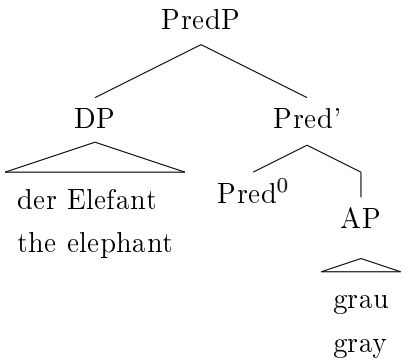
The functional category D contains the nominal agreement features case, number and gender. Because it is not finally settled whether functional projections related to these features exist, the parse tree in (2.7) does not contain these projections.

Adjectives have also been analyzed as NP-adjuncts or as adjoined to other possible functional projections inside the DP. As a consequence, multiple APs inside the DP must be analyzed as recursive conjunction to the same category. This analysis cannot account for the ordering restrictions on multiple adjectives, which I describe later in this section. Analyzing adjectives as specifiers of functional heads in the extended projection of N has the advantage that APs stand in a specifier-head relation with the head of the NP. This can account for the agreement of attributive adjectives and nouns (Alexiadou et al., 2007). Theories disagree on whether adjectives should be analyzed as heads or as maximal projections. For an overview of the advantages and disadvantages of both analyses see Alexiadou et al. (2007). The fact that German adjectives can take complements, e.g., *die* ***auf ihren Sohn*** *stolze Mutter* (*'the proud of her son mother'), suggests an analysis as maximal projections. If adjectives were heads, the prenominal adjective would not be able to take the prepositional phrase as a complement because it already has a phrasal projection – the NP – as its complement (Alexiadou et al., 2007).

Predicative structures are structures in which the adjective is the complement of the copula. These structures have been usually analyzed as Small Clauses (Stowell, 1978, and subsequent work). Bowers (1993) introduces a dedicated functional projection for predicative structures, PredP, which constitutes the Small Clause. This additional projection is only required for predicative structures, but not for attributive ones. The Pred-head selects a XP-complement, e.g., an AP. The subject of the predicative sentence is in the specifier of PredP (see example (2.8)). Hence, subject DP and adjective form one constituent.

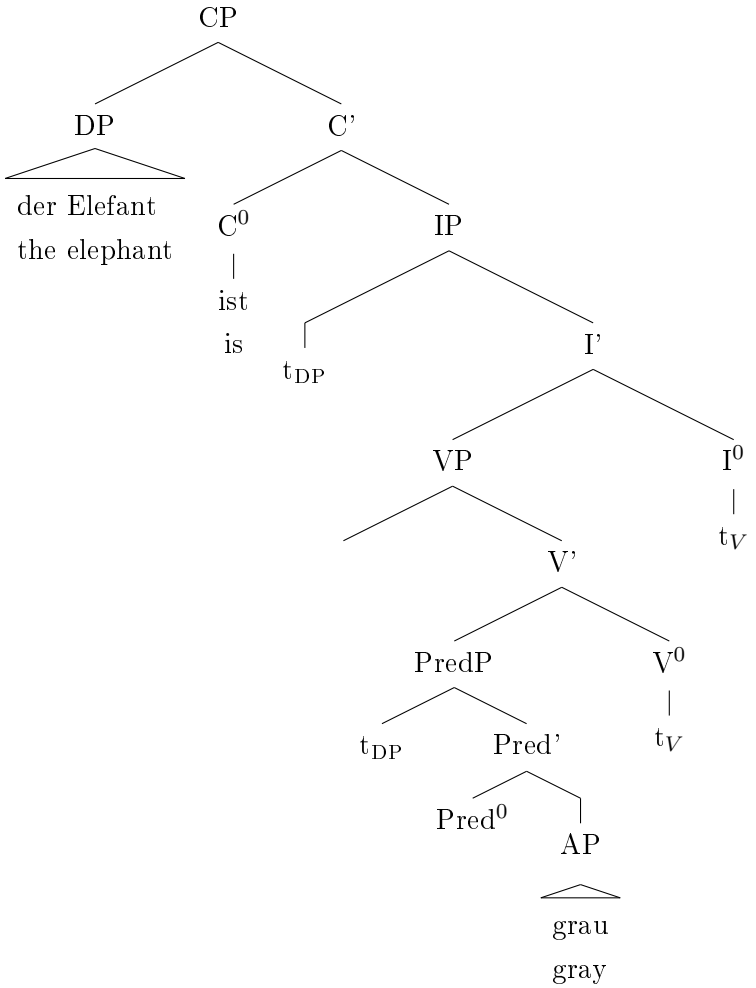
(2.8) Der Elefant ist grau.

'The elephant is gray.'



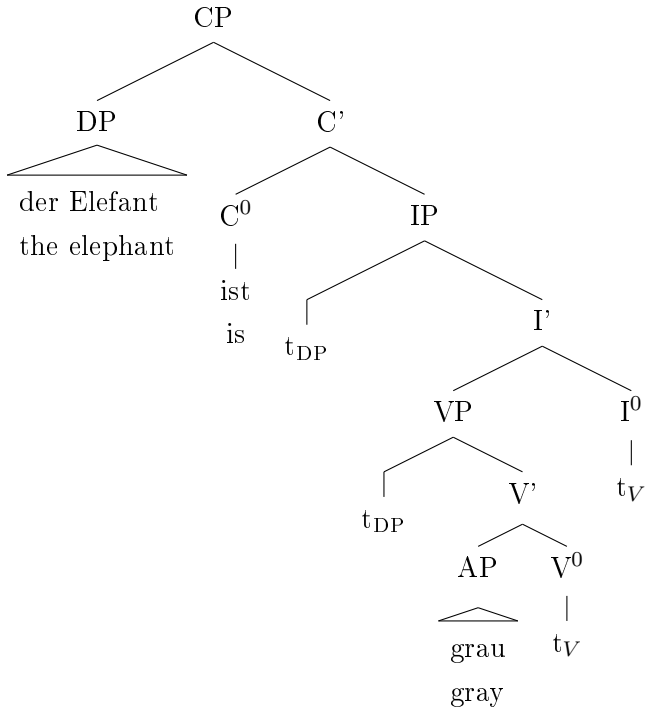
The copula is not part of the constituent because it is often assumed to not contribute any semantic content, but only to express finiteness. Bowers' analysis of predicative structures (see (2.5)) is similar to den Dikken's (2006). However, den Dikken's analysis differs in the abstractness of the RELATOR-head and the non-directionality of predicative relationships, which results in the same structure for predicative and attributive structures. There are different accounts on the syntactic analysis of the copula itself (see Maienborn, 2003). The copula can be for example the phonetic realization of Pred^0 as in (2.8) or R^0 in (2.5). In (2.9) the copula is a raising verb that selects a Small Clause complement:

(2.9) Der Elefant ist grau.
'The elephant is gray.'



Maienborn (2003) analyzes the copula as an ordinary verb that selects a predicative XP-complement as displayed in (2.10).

(2.10) Der Elefant ist grau.
'The elephant is gray.'



Maienborn (2003) argues that this allows for the fact that it is the copula which establishes the predicative relation between the subject and the complement, in this case the adjective. The question how exactly the copula establishes the relation between subject and adjective has been answered differently. In line with Davidson (1967), it has been argued that copula constructions denote situations or states, in which the property expressed by the adjective holds of the subject referent. Hence, copula constructions have been analyzed like other (stative) verbs. The question is whether it is the copula itself or the adjective which introduces the situation argument. If the former is the case, the copula makes a semantic contribution. In this approach, *sein* ('be') is analyzed similarly to other main verbs, but instead of denoting a property it introduces an argument (see Bierwisch, 1988; Kamp & Reyle, 1993; Rothstein, 1999). The copula and its complement, i.e., the adjective, denote a situation. The crucial contribution of the copula is the introduction of a referential situation argument, which is specified by the adjective. A disadvantage of this kind of analysis is that the theta-role of the subject of the copular construction is difficult to define. However, if copulas behave similarly to other main verbs, it should be possible to assign a theta-role to the subject.

Other researchers (e.g., Kratzer, 1995) claim that the adjective introduces the situation argument. In this analysis predicates are divided into stage-level and individual-level predicates. Only stage-level predicates refer to situations, and hence introduce a situation argument. This is mirrored for example in the distribution of the Spanish copula forms *ser* and *estar*. One form selects for stage-level predicates, the other form selects for

individual-level predicates. The German copula *sein* is ambiguous; it selects both stage-level and individual-level predicates. The different readings can for instance be achieved by functional projections which are present for one type of predicate but not for the other (Kratzer, 1995). The disadvantage of this analysis is that the classification into stage-level and individual-level predicates is not always clear-cut, in particular with respect to adjectives. *Funny* for example passes the typical tests for individual-level predicates: it has only the generic reading with bare plurals (e.g., *Comedians are funny*), it is not possible in *there*-constructions (e.g., **There are comedians funny*), and they are not compatible with perception verbs (e.g., *Mary saw John funny*). But *funny* can occur in *when*-clauses (e.g., *When Dumbo is funny, his audience laughs*), which should only be possible for stage-level predicates.

Maienborn (2003) argues against a reference to situations. According to her analysis copular constructions denote so-called *K-Zustände* ('K-states').[3] *K-Zustände* are abstract objects (a subtype of facts) for the exemplification of a property P on an entity x at a time t (Maienborn, 2003, p. 121). The copula's contribution is to refer to these K-states. Independent of the theory and independent of the assumptions whether the copula has semantic content, it can be concluded that the copula has a designated function: it establishes a relation between the subject and the property denoted by the adjective.

In summary, the relation between attributive and predicative structures is not finally resolved. Because the focus of this thesis is on the semantics of adjectives, I address syntactic questions when they are relevant for the semantics of adjectives in the remainder of this thesis.

In the examples so far, only a single adjective occurred prenominally inside the DP or as the complement of the copula, but multiple adjectives are possible as well. In predicative position, multiple adjectives can be ordered freely as in (2.11a) and (2.11b). In contrast, in attributive position the adjective order is subject to restrictions as in (2.11c) versus (2.11d) (e.g., Sproat & Shih, 1991).

(2.11) a. Der Kaffee ist fantastisch, groß, heiß, schwarz.
'The coffee is fantastic, tall, hot, black.'

b. Der Kaffe ist groß, schwarz, heiß, fantastisch.
'The coffee is tall, black, hot, fantastic.'

c. Ein fantastischer, großer, heißer, schwarzer Kaffee
'A fantastic, tall, hot, black coffee'

d. *Ein schwarzer, fantastischer, großer, heißer Kaffee
*'A black, fantastic, tall, hot coffee'

The observed canonical order of adnominal adjectives is the following (Laenzlinger, 2005):

[3] The term *K-Zustand* refers to Kim (1969, 1976) cited by Maienborn (2003).

(2.12) [QUANTITY Ordinal > Cardinal] > [SPEAKER-ORIENTED Subjective Comment > Evidential] > [INTERNAL PHYSICAL PROPERTY Size > Length > Height > Speed > Depth > Width] > [MEASURE Weight > Temperature > Wetness > Age]
> [EXTERNAL PHYSICAL PROPERTY Shape > Color > Nationality > Material]

It is a controversial question whether this ordering is regulated by syntactic structure, i.e., a rigid order of functional projections, or whether it is semantically driven. This topic is not further addressed in the remainder of this thesis.

This section described the two discourse functions of adjectives and their corresponding syntactic structures. However, the exact nature of their derivations and the relationship between attributive and predicative structures are still topics in syntactic research. The challenge is that both structural and semantic components play a role and are often intermixed.

Section 2.2 is devoted to the adjectives' meaning.

2.2 Semantics of adjectives

In this part of the chapter, section 2.2.1 takes up again the adjective classes mentioned in (2.12). Section 2.2.2 examines whether the adjectives' different functions (and syntactic positions) must be reflected by their semantic analysis. Section 2.2.3 refines the semantic properties figured out in Section 2.2.2.

2.2.1 Notional properties

Ordering restrictions for multiple prenominal adjectives as in (2.12) are often described by means of so-called notionally-based adjective classes, a term adopted from McNally (2016). Notionally-based categories have been mentioned first by Dixon (1982). For languages with a word class ADJECTIVE, he showed that the semantic content, that is the concepts expressed by adjectives, is similar across these languages. Isolating the concepts[4] associated with the ADJECTIVE class, he suggests seven semantic types as universal concepts expressed by adjectives: DIMENSION, PHYSICAL PROPERTY, COLOR, HUMAN PROPENSITY, AGE, VALUE, and SPEED. In subsequent work, the classes introduced by Dixon (1982) have been further specified into subcategories. Blackwell (2005) based her classification on Dixon (1982) and on the specifications by Frawley (1992) and Roget (1965). Blackwell's specifications concern in particular the PHYSICAL PROPERTY class, which contains the subcategories Texture, Configuration, Substantiality, Sense, Consistency, Matter, Form, Speed, Wetness, Cleanliness, Appearance, and Edibility. In Blackwell (2005), the HUMAN PROPENSITY class distinguishes between adjectives expressing Mental State, Physical State and Behavior. An additional class subsumes all other adjectives, expressing for instance Endearment, Similarity, Sufficiency, and Conformity. Tribushinina et al. (2014) used a modified version of Blackwell's classification. They differentiate the following

[4] Dixon calls them semantic types.

classes: Age, Behavioral Property, Color, Conformity, Evaluation, Internal State, Modal, Ordinal Number, Physical State, Quantitative Characteristics, Spatial Property, Temporal Property, Other and Physical Property with Surface, Configuration/Functionality, Taste/Edibility, Smell, Sound, Shape, Light, Consistency, Matter, Speed, and Temperature as subcategories.

According to McNally (2016), notionally-based typologies classify modifiers according to their descriptive content. Hence, notionally-based classes tell more about the adjectives themselves than about their function as modifiers. Thus, they are often used to account for phenomena that do not necessarily involve modification such as the inventory of adjectives across languages (Dixon, 1982) or the composition of the mental lexicon (Blackwell, 2005; Tribushinina et al., 2014). In addition, they are often found in descriptive grammars and are used to depict distributional patterns such as the adjective ordering restrictions mentioned in (2.12), positional restrictions (attributive vs. predicative), or combinations with specific morphemes (McNally, 2016).

One shortcoming of notionally-based typologies is that diverse classification schemes and terminologies exist. Moreover, it is difficult to determine how many adjective categories are necessary (McNally, 2016; Morzycki, 2016), as it is already apparent by the classifications decribed above. Another shortcoming of notionally-based typologies is illustrated by example (2.13) taken from Blackwell (2005, p. 541):

(2.13) a. I don't think you're quite strong enough to blow a tree down.
b. That balloon is strong.

In (2.13a), *strong* modifies an animate referent, thus it should be coded as Human Propensity. In (2.13b), *strong* modifies an inanimate referent, hence it should be coded as Physical Property. Therefore, the same adjective can belong to different classes depending on the properties of the modified referent. This 'ambiguity' prohibits a uniform semantic description of the adjective's properties. Tribushinina et al. (2014) mention another issue of one particular class, the Physical Property class, with respect to language acquisition research. Because this class is very large and very diverse, it may be difficult to detect developmental patterns as they may be disguised by the heterogeneity of the class.

Section 2.2.2 focuses on so-called entailment-based typologies (McNally, 2016), which are more prominent in formal semantics because they focus on the adjective's function as nominal modifier.

2.2.2 Entailment properties

Entailment-based typologies focus on the adjective's function as a modifier and classify adjectives according to the inferences they license (McNally, 2016). As described in Section 2.1, adjectives can occur either in prenominal position (attributive) or as complement of a copula (predicative). Depending on their syntactic position, adjectives exhibit different discourse functions. In predicative position adjectives describe a property of an entity. In

(2.14c) this is the subject NP. Hence, the adjective specifies the entity itself rather than specifying the set of possible referents. Therefore, predicative adjectives can be interpreted as properties or one-place-predicates of type $\langle e, t\rangle$ similar to other predicates such as nouns and verbs (Demonte, 2011; Hamann, 1991). This is exemplified in (2.14). x is a variable over individuals.

(2.14) a. Dumbo fliegt. λx.(**fliegen**(x))
'Dumbo flies.'

b. Dumbo ist ein Elefant. λx.(**Elefant**(x))
'Dumbo is an elephant.'

c. Dumbo ist männlich. λx.(**männlich**(x))
'Dumbo is male.'

The examples show that the lexical entries of the respective predicates have the same form. In contrast to predicative adjectives, attributive adjectives modify the kind denoted by the noun, which is a set of entities, e.g., elephants as in *der männliche Elefant* ('the male elephant'). The adjective specifies the entities to be considered, and thus restricts the set of possible referents. Therefore, attributive adjectives are often analyzed as modifiers, i.e., as functions from properties to properties (type $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$) as illustrated in (2.15). P is a variable for a property of individuals.

(2.15) $[\![\textit{männlich}]\!]$=$\lambda$P.$\lambda$x.**männlich**(x)$\wedge$ P(x)

If the adjective is analyzed as a modifier, it combines with the noun via *Functional Application* (see (2.16)), which is illustrated in (2.17).

(2.16) Functional Application (Heim & Kratzer, 1998, p. 44)
If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters , and $[\![\beta]\!]$ is a function whose domain contains $[\![\gamma]\!]$, then $[\![\alpha]\!] = [\![\beta]\!]([\![\gamma]\!])$.

(2.17) $[\![\textit{männlicher Elefant}]\!]$=$[\![\textit{männlich}]\!]$ ($[\![\textit{Elefant}]\!]$)
= λx.$[\![\textit{männlich}]\!](x)\wedge$ $[\![\textit{Elefant}]\!]$(x)
= λx.**männlich**(x)$\wedge$ **Elefant**(x)

Semanticists disagree on whether adjectives used as prenominal modifiers denote functions from properties to properties or whether they can be analyzed as denoting properties of individuals similar to (2.14c). Because all adjectives belong to the same word class, it would be possible that they all have the same semantic type. This analysis can be called Adjective Type Homogeneity Hypothesis (see Morzycki, 2016). However, as I show in this section, different adjectives exhibit different relations between adjective and noun indicating that they may also be analyzed differently. This analysis can be called Adjective Type Heterogeneity Hypothesis (see Morzycki, 2016). These two hypotheses have already been discussed by Kamp (1975). Assuming Adjective Type Homogeneity all adjectives denote modifiers of type $\langle\langle s, \langle e, t\rangle\rangle\langle e, t\rangle\rangle$. Because this is the only analysis

possible for some adjectives such as *ehemalig* ('former'), it is applied to all other adjectives ('generalizing to the worst case', Partee and Hendriks (1997) in the spirit of Montague (1970)). However, there are also attempts to reduce all adjectives to type $\langle e, t\rangle$. In contrast, under the ADJECTIVE TYPE HETEROGENEITY HYPOTHESIS, some adjectives denote properties of type $\langle e, t\rangle$ and some adjectives are modifiers of type $\langle\langle s, \langle e, t\rangle\rangle\langle e, t\rangle\rangle$. Importantly, both hypotheses make the same predictions about the truth conditions of attributive adjectives.

In order to see what is controversial about the semantic analysis of attributive adjectives, a closer look at the relationship between adjective and noun is necessary. The relation between adjective and noun is captured by entailment-based properties of adjectives. Depending on these properties different types of inferences are licensed. The two main properties are the following:

(2.18) Subsectivity

$$\frac{\text{X is a Adj N.}}{\text{X is a N.}}$$

(2.19) Intersectivity

$$\frac{\text{X is a Adj N.}}{\text{X is Adj and X is a N.}}$$

In (2.18), the sentence above the single line is the premise and the sentence below the single line is the conclusion. In other words, the sentence below the single line can be inferred from the sentence above the single line. In (2.19), the double line indicates that the inference holds in both directions: if *X is a Adj N* then *X is Adj and X is a N* and if *X is Adj and X is a N* then *X is a Adj N*. This scheme requires that a (synonymous) predicative counterpart exists for the prenominal adjective. However, some intersective attributive adjectives cannot be used predicatively or the predicative use results in a different interpretation. The scheme in (2.20) avoids the requirement for a predicative counterpart, but illustrates intersectivity equivalently (Ede Zimmermann, p.c.).

(2.20)

$$\frac{\text{X is a Adj N1.}\quad \text{X is a N2.}}{\text{X is a N1 and X is a Adj N2.}}$$

In principle, subsectivity and intersectivity are independent properties of adjectives. However, the three combinations of subsectivity and intersectivity in (2.21) are attested in the literature on the semantics of adjectives (e.g., Kamp & Partee, 1995):

(2.21)
a. Adjectives that are subsective and intersective.
b. Adjectives that are subsective, but not intersective.
c. Adjectives that are neither subsective nor intersective.

Note that the term 'subsective' is sometimes used to refer to non-intersective adjectives (e.g., Kamp & Partee, 1995; Partee, 1995). In the remainder of this thesis I use 'subsective'

in its literal sense, namely for all adjectives for which the inference in (2.18) holds: The set denoted by the adjective-noun phrase is a subset of the set denoted by the noun. Adjectives for which intersectivity does not apply are called 'non-intersective'.

The following examples illustrate the attested combinations of subsectivity and intersectivity in (2.21) by means of specific adjectives. The adjective *männlich* ('male') is subsective and intersective. The sentence in (2.22) implies that Dumbo is an elephant.

(2.22) Dumbo ist ein männlicher Elefant.
'Dumbo is a male elephant.'

Männlich ('male') is a subsective adjective because the set of male elephants is a subset of the set of elephants.

(2.23) $[\![\textit{männlicher Elefant}]\!] \subseteq [\![\textit{Elefant}]\!]$

Moreover, one can infer from (2.22) that (i) Dumbo is male and that (ii) Dumbo is an elephant. Likewise, if one knows that (i) and (ii) are true of an entity, one can infer that the sentence in (2.22) is also true of that entity. The meaning of the complex adjective-noun phrase follows from the conjunction of the two properties denoted by the adjective and by the noun. This mode of composition is called *Predicate Modification* and follows the rule in (2.24):

(2.24) Predicate Modification (Heim & Kratzer, 1998, p. 65)
If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters , and $[\![\beta]\!]$ and $[\![\gamma]\!]$ are both in $\mathrm{D}_{<e,t>}$, then
$[\![\alpha]\!] = \lambda \mathrm{x} \in \mathrm{D}_e.\ [\![\beta]\!](\mathrm{x}) = [\![\gamma]\!](\mathrm{x}) = 1.$

In (2.25), this rule is applied to the sentence in (2.22):

(2.25) $[\![\textit{männlicher Elefant}]\!] = \lambda \mathrm{x}.[\![\textit{männlich}]\!](\mathrm{x}) = [\![\textit{Elefant}]\!](\mathrm{x}) = 1$
$= \lambda \mathrm{x}.\mathbf{männlich}(\mathrm{x}) = \mathbf{Elefant}(\mathrm{x}) = 1$
$= \lambda \mathrm{x}.$ x ist männlich $\wedge$ x ist ein Elefant

In order to apply Predicate Modification as the mode of composition, both adjective and noun must denote properties of individuals (type $\langle e, t\rangle$) as in (2.14b) and (2.14c). In set theoretical terms, they denote sets of individuals. The meaning of the adjective-noun phrase in (2.22) is the intersection of the set denoted by the adjective and the set denoted by the noun (e.g., Kamp & Partee, 1995; Partee, 1995).

(2.26) $[\![\textit{männlicher Elefant}]\!] = [\![\textit{männlich}]\!] \cap [\![\textit{Elefant}]\!]$

This shows that intersective adjectives like *männlich* ('male') can be analyzed as properties of individuals and combine with a noun via Predicate Modification. Hence, they need not to be analyzed as modifiers of type $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$ and need not to combine with the noun via functional application as in (2.15).

Intersective adjectives have an invariant meaning, i.e., the meaning is independent of the modified noun. As a consequence, the inference in (2.20) is possible. This is illustrated in (2.27) for *männlich* ('male').

(2.27) Dumbo ist ein männlicher Elefant. ('Dumbo is a male elephant.')
Dumbo ist ein Zirkusartist. ('Dumbo is a circus artist.')

Dumbo ist ein Elefant und ein männlicher Zirkusartist.
('Dumbo is an elephant and a male circus artist.')

Adjectives like *klein* ('small') or *groß* ('big') are subsective just like *männlich* ('male'). Small elephants are a subset of elephants, and big mice are a subset of mice. But in contrast to adjectives like *männlich* ('male'), the meaning of adjectives such as *klein* ('small') or *groß* ('big') varies depending on the head noun. Comparing the examples in (2.28) shows that it is not enough to operate only with the set of mice, the set of elephants, the set of small things, and the set of big things.

(2.28) a. Dumbo ist ein kleiner Elefant.
'Dumbo is a small elephant.'
b. Mickey ist eine große Maus.
'Mickey is a big mouse.'

In these examples, Dumbo for instance must be in the set of small things, but also in the set of big things because he is an elephant, and hence bigger than mice. The same holds for Mickey in the opposite direction (see Morzycki, 2016). Thus, it is not possible to infer from *Dumbo ist ein kleiner Elefant* ('Dumbo is a small elephant') that he is small. At first sight, it seems that the Predicate Modification rule cannot apply to adjectives like *klein* ('small'). Consequently, they should be treated as non-intersective. Intuitively, intersection is not possible for this sort of adjectives because the extension of the adjective itself is not a set. The adjective must be interpreted relative to the extension of the noun. The noun serves as the comparison class. Hence, *Dumbo ist klein* ('Dumbo is small') must be interpreted as *Dumbo ist klein für einen Elefanten* ('Dumbo is small for an elephant'). I discuss the role of the comparison class for the interpretation of adjectives such as *klein* ('small') in Section 2.2.3. For now, the crucial observation is that intersectivity can be fixed if one assumes that the adjective takes the comparison class C as an additional argument as suggested by Morzycki (2016) and others.

(2.29) $[\![kleiner\ Elefant]\!]=\lambda x.[\![klein]\!](x)(C)\wedge [\![Elefant]\!](x)$

In example (2.29), the comparison class equals the set denoted by the noun.

An adjective that has been analyzed as subsective but non-intersective is *begabt* ('skillful'), for which the inference in (2.30) is not possible (Kamp & Partee, 1995).

(2.30) Dumbo ist ein begabter Elefant. ('Dumbo is a skillful elephant.')
Dumbo ist ein Zirkusartist. ('Dumbo is a circus artist.')

Dumbo ist ein Elefant und ein begabter Zirkusartist.
('Dumbo is an elephant and a skillful circus artist.')

Morzycki (2016) points out that fixing the comparison class is not sufficient for this kind of modification. Although possible, it cannot explain why a skillful circus artist is someone who for instance fire-breathes skillfully. Instead of saying that someone is *begabt für einen Zirkusartisten* ('skillful for a circus artist') one would rather say that someone is *begabt als ein Zirkusartist* ('sklillful as a circus artist'). The paraphrase points out two things: first, this class of adjectives cannot single out the set of individuals who have the property of for instance just being skillful (Partee, 1995). Second, this kind of adjective is dependent on the intension of the modified noun rather than on its extension as in the previous example. Hence, the intersection of the set denoted by the noun and the set denoted by the adjective is not possible because neither the adjective nor the noun refer to sets. Therefore, the Predicate Modification rule cannot apply. As a consequence, adjectives of this class do not denote properties of individuals, but functions from properties to properties (type $\langle s, \langle e, t\rangle\rangle, \langle e, t\rangle\rangle$). Thus, the lexical entry of *begabt* ('skillful') is the following (see Morzycki, 2016).

(2.31) $[\![\textit{begabt}]\!]^w = \lambda \mathrm{P}.\lambda \mathrm{x}.\textbf{begabt-als}(\mathrm{P})(\mathrm{x})(\mathrm{w})$

Similar to the adjectives mentioned so far, for *begabt* ('skillful') the extension of the modified noun is a subset of the extension of the noun.
However, there exist approaches that apply an intersective analysis to adjectives like *begabt* ('skillful') (e.g., Larson, 1999; von Fintel & Heim, 1999). Larson (1999) argues that the distinction between intersective and non-intersective readings should be separated from the distinction between adjectives denoting properties and adjectives denoting functions from properties to properties. He claims that both intersective and non-intersective adjectives are of type $\langle e, t\rangle$. He suggests that non-intersective readings of adjectives like *begabt* have a different source: an event argument e (type v). This analysis is similar to the analysis of manner adverbs. Larson (1999) analyzes manner adverbs as predicates of Davidsonian events and assigns to them an intersective, property-denoting interpretation. In the following, I describe Larson's analysis of adjectives such as *begabt* ('skillful') in Morzycki's (2019) slightly modified version.[5] The first assumption is that nouns, e.g., fire-breather, can be understood in terms of events. The fire-breathing event is expressed by a generic quantifier GEN. In (2.32), "the generic event among the ones relevant in the discourse context c is [a fire-breathing by Dumbo]" (Morzycki, 2016, p. 37).

[5] Morzycki (2016) notes that he implemented Larson's idea in a simplified framework, but that Larson's core analysis is preserved.

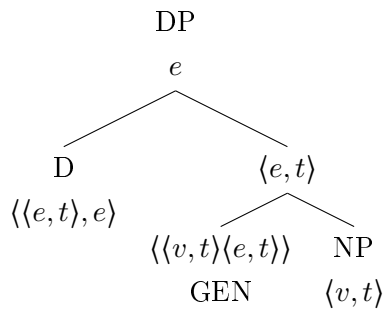
(2.32) Dumbo spuckt Feuer. ('Dumbo fire-breathes.')
GEN e: $\mathbf{relevant}_c\ (e)\ [\mathbf{feuerspucken}(e)(\mathbf{Dumbo})]$

By treating *Feuerspucker* ('fire-breather') as a property of fire-breathing events, this is incorporated in the denotation of the noun:

(2.33) $[\![Feuerspucker]\!] = \lambda e.\mathbf{feuerspucken}(e)$

However, this denotation is not a property of individuals, hence not of the right type to occur in the nominal position of a DP. Morzycki (2016, p. 38) suggests to introduce a generic quantifier as a node in the tree as in (2.34). Applied to a property of events, this generic quantifier yields a property of individuals. As a result, a fire-breather is the agent of a typical fire-breathing event.

(2.34)



If an adjective is inserted, it can either apply to events, yielding a non-intersective reading, or to individuals, yielding an intersective reading. In non-intersective readings, the adjective occurs below GEN and can be interpreted intersectively with the noun. Following Morzycki (2016), a skillful fire-breather is the agent of the typical event that is skillful and fire-breathing. In contrast, in the intersective reading the adjective occurs above GEN. In this reading, a skillful fire-breather is someone who is skillful and a fire-breather. The crucial point is that in both readings adjective and noun are comined intersectively. As noted by Morzycki (2016), the advantage of this kind of analysis is that non-intersective readings do not result from peculiarities of some adjectives, but from aspects of the structure of the extended nominal projection. However, a shortcoming of the analysis is the introduction of event arguments for nouns where it is less intuitive, i.e., nouns that are not deverbal or inanimate. Both Larson (1999) and Morzycki (2016) use the famous *beautiful dancer*-example; the example used here (*skillful fire-breather*) also seems compatible with the analysis. For nouns such as *friend* or *dress* it is unclear whether adjectives like *beautiful* have the same readings and whether they can be analyzed in the same way.

The 'implicit argument approach' is another approach aiming at reducing non-intersective to intersective readings (see e.g., von Fintel & Heim, 1999). According to this approach, non-intersective adjectives have an additional argument slot, which is saturated in their extended projection with an *as*-phrase. In attributive uses, the *as*-phrase is implicit. Again, adjective and noun are interpreted intersectively as shown in (2.35) (adapted from Morzycki, 2016, p. 41).

(2.35) $[\![\textit{begabt}\ \cancel{\textit{als Feuerspucker}}\ \textit{Feuerspucker}]\!]$
$= \lambda x.\ [\![\textit{begabt}\ \cancel{\textit{als Feuerspucker}}]\!] \wedge [\![\textit{Feuerspucker}]\!](x)$
$= \lambda x.\ \textbf{begabt-als}(\textbf{Feuerspucker}) \wedge \textbf{Feuerspucker}(x)$

In this approach, the intensionality of adjectives like *begabt* ('skillful') becomes visible by the (implicit) *as*-phrase, but it does not affect the compositional mechanism, i.e., it does not prevent the intersection of the adjective and the noun denotation.

The three adjective classes introduced so far are subsective. Some adjectives, however, are neither subsective nor intersective. On the one hand, this concerns modal adjectives like *angeblich* ('alleged') as in (2.36a). They are non-intersective because no set of alleged entities exists. They are non-subsective because the set of alleged circus artists is not a subset of the set of circus artists: it can contain individuals for which it is not true that they are or will become a circus artist. On the other hand, non-subsectivity also applies to temporal adjectives like *ehemalig* ('former'): a former circus artist is not in the set of circus artists at the time a sentence like (2.36b) is uttered. Just like for *angeblich* ('alleged'), it is not possible to define a set of former entities.

(2.36) a. Dumbo ist ein angeblicher Zirkusartist.
'Dumbo is an alleged circus artist.'
b. Dumbo ist ein ehemaliger Zirkusartist.
'Dumbo is a former circus artist.'

Because *angeblich* ('alleged') quantifies over possible worlds w and *ehemalig* ('former') over times s, an intensional semantics is called for. These adjectives must be analyzed as functions that apply to noun meanings (type $\langle\langle s, \langle e, t\rangle\rangle\langle e, t\rangle\rangle$) as in (2.37) (see Morzycki, 2016, p. 24; 46).

(2.37) a. $[\![\textit{angeblich}]\!] = \lambda P.\ \lambda x.\lambda w.\ \forall\ w' \in \textbf{allegiations}(w)[P(x)(w')]$
b. $[\![\textit{ehemalig}]\!] = \lambda P.\lambda x.\lambda s.\ \exists\ s'.\ [s' \prec_{time} s \wedge P(x)(s') \wedge\neg\ P(x)(s)]$

The adjective takes the noun as an argument and therefore adjectives from this class cannot occur in predicative position. Note that the sentence in (2.36a) can be paraphrased as *Angeblich ist Dumbo ein Zirkusartist* ('Allegedly, Dumbo is a circus artist') and the sentence in (2.36b) can be paraphrased as *Früher war Dumbo ein Zirkusartist* ('In former times, Dumbo has been a circus artist'). Hence, alternatively to the modifier analysis non-subsective/non-intersective adjectives can be analyzed as sentence adverbs (type $\langle\langle s, t\rangle, t\rangle$).

The adverbial reading is also available for adjectives such as *occasional*. A sentence like *An occasional sailor strolled by* can receive the so-called 'nonlocal' reading *Occasionally, a sailor strolled by* (Bolinger, 1967): the adjective does not interact with the adjacent noun, but has a wider scope. Possibly, non-subsective/non-intersective adjectives belong to the class of adjectives with nonlocal readings and receive an entirely different analysis than the subsective/intersective adjectives with a local interpretation. Because it is still open how the nonlocal reading is explained best (see Morzycki (to appear) for an overview

of analyses and a recent proposal), I assume the more common modifier-analysis of non-subsective/non-intersective adjectives.

This overview over the entailment-based properties of adjectives and their distribution in different kinds of adjectives revealed the controversy with regard to (attributive) adjectives. Because they all belong to the same word class, it would be possible that all adjectives have the same semantic type (Adjective Type Homogeneity Hypothesis). According to this view, all adjectives, including intersective ones, denote modifiers of type $\langle\langle s, \langle e, t\rangle\rangle\langle e, t\rangle\rangle$. However as mentioned above, there are also attempts to reduce all adjectives to intersective ones of type $\langle e, t\rangle$. However, as shown in this section, different adjectives exhibit different relations between adjective and noun indicating that they may also be analyzed differently (Adjective Type Heterogeneity Hypothesis). In this approach, intersective adjectives denote properties of type $\langle e, t\rangle$ and non-subsective adjectives are of a more complex semantic type like $\langle\langle s, \langle e, t\rangle\rangle\langle e, t\rangle\rangle$.

Acquisition data may contribute to theory building by investigating whether different kinds of adjectives are acquired differently. In particular, showing that different adjectives are acquired at different ages or that one interpretation is generalized across adjectives may reveal insights about the underlying semantic structure of the respective adjectives and the compositional mechanisms of adjective-noun phrases. At the same time, predictions about the acquisition of adjectives are guided by formal semantic considerations. I come back to that in Section 2.4.

For the time being, I assume type heterogeneity. This means that not all attributive adjectives must be analyzed as modifiers. As a consequence, the syntactic positions and their corresponding discourse functions/variations in meaning described in section 2.1 cannot be mapped one-to-one to compositional differences. This lacking one-to-one correspondence may point to an analysis similar to den Dikken's, which assumes the same syntactic configuration for attributive and predicative adjectives. Nevertheless, for the time being it is unclear how syntax and semantics are related exactly.

In (2.21) three adjective classes were suggested:

a. Adjectives that are subsective and intersective.

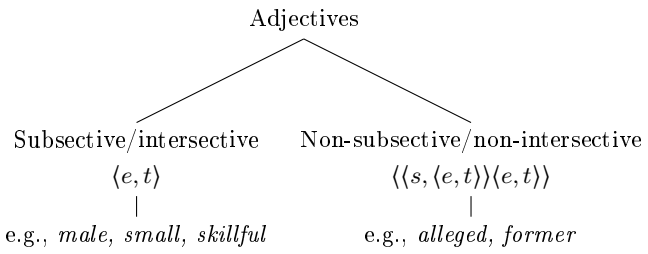
b. Adjectives that are subsective, but not intersective.

c. Adjectives that are neither subsective nor intersective.

Rather than these three classes, I assume two classes of adjectives: subsective/intersective adjectives as in a. (e.g., *male, small, skillful*) and non-subsective/non-intersective adjectives as in c. (e.g., *alledged, former*). I do not assume a class of subsective/non-intersective adjectives as in b. for two reasons. First, Morzycki (2016) notes that apparently non-intersective adjectives "seem to occupy a middle ground between intersective and non-subsective, more complicated than the former but not as complicated as the latter. But as far as the types themselves are concerned, there is no middle ground to occupy. You denote either a property or predicate modifier." (Morzycki, 2016, p. 27). One could argue that in terms of types

the distinction is between extensional (e.g., *male, small*) and intensional (e.g., *skillful, alleged*) modifiers. However, and this is the second reason, it has been shown in this section that apparently non-intersective adjectives such as *small* and *skillful* can receive an intersective and extensional analysis, hence it is possible to assign them an extension of type $\langle e, t\rangle$. Based on the entaillment-based properties the following classification of adjectives in (2.38) emerges:

(2.38)



In order to capture the differences within the class of subsective/intersective adjectives, I refine this class by adding another important property of adjectives: gradability.

2.2.3 Gradability properties

Differences within the class of adjectives that I refer to as subsective/intersective can be captured by their gradability properties. Gradability is another important property of adjectives besides their entailment properties. Gradability or comparison is often used as a criterion to distinguish adjectives from other word classes such as nouns and verbs (e.g., Dixon, 2004). In addition, Kennedy (2006) points out that ordering and comparing objects is a basic component of human cognition. Natural languages reflect this cognitive ability with syntactic categories that express gradable concepts.

Not all adjectives are gradable. Hence, the first distinction to make is between non-gradable and gradable adjectives. Gradable adjectives like *klein* ('small') or *begabt* ('skillful') can occur in comparative constructions (see example (2.39)) and can be combined with degree modifiers (see example (2.40)), whereas non-gradable adjectives such as *männlich* ('male') normally cannot.

(2.39) a. Mickey ist kleiner als Dumbo.
'Mickey is smaller than Dumbo.'

b. Dumbo ist begabter als Mickey.
'Dumbo is more skillful than Mickey.'

c. ? Dumbo ist männlicher als Mickey.
'Dumbo is more male than Mickey.'

(2.40) a. Mickey ist zu klein.
'Mickey is too small.'

b. Dumbo ist sehr begabt.
'Dumbo is very skillful.'

c. ? Dumbo ist sehr/zu männlich.
'Dumbo is very/too male.'

Bierwisch (1987, 1989) notes that combining adjectives like *männlich* ('male') with comparative morphemes and degree modifers is in general possible. However, Bierwisch among others classify *male* as non-gradable.[6] Other non-gradable adjectives are *verheiratet* ('married'), *deutsch* ('German') as well as color adjectives. With regard to color adjectives, Kennedy and McNally (2010) claim that they have a gradable and a non-gradable reading. The gradable reading refers to either the color quantity or the color quality. The non-gradable reading means the classificatory reading as in *red wine* for instance. Demonte (2011) argues that although color adjectives can accept degree modifiers and can occur with comparative morphology, they differ from gradable adjectives with respect to the content of their scale: scales of gradable adjectives consist of degrees whereas scales of color adjectives consists of prototypes and shades of the respective color. Hansen and Chemla (2017) found in their experimental study that color adjectives are context-sensitive (e.g., red with respect to wine versus red with respect to blood), but that the so-called standard of comparison is not contextually variable unlike that for adjectives such as *small*. I assume color adjectives to be non-gradable. An additional argument for their non-gradability is that they do not come in antonym pairs which is common for gradable adjectives. Non-subsective/non-intersective adjectives are not gradable either. In the following, when I refer to non-gradable adjectives, it is restricted to subsective/intersective adjectives because I use gradability as a property to refine this class.

It is widely assumed that gradable adjectives differ from non-gradable adjectives in that they denote relations between entities and degrees (Cresswell, 1976; von Stechow, 1984; Kennedy, 2007, among others). Degrees are understood as representations of measurement. A set of degrees that is totally ordered with respect to some dimension (e.g., HEIGHT) is called a scale. A measure function (type $\langle e, d \rangle$) assigns a degree to an individual. For instance, the measure function for height ($\lambda x.x's\ height$) applied to an entity x yields the (unique maximal) degree to which x is tall. The measure function on the scale of height can be written as μ_{tall}. Approaches differ with regard to the nature of degrees (for an overview see Solt & Gotzner, 2012). The so-called 'derived degree approach' assumes that degrees and scales are derived from ordering relations between individuals that result in equivalence classes (Cresswell, 1976). In contrast, the so-called 'abstract degree approach' assumes that degrees are primitives and independent of specific measurement relations between entities (von Stechow, 1984; Kennedy, 2007). von Stechow (1984) argues that sentences such as *Ede is taller than he is broad* can only be analyzed by assuming abstract degrees because two measures of different scales (HEIGHT and BREADTH) of one individual are compared. It is not the case that the individual is compared to itself and that this results in equivalence classes from which a scale is derived. Ede Zimmermann (p.c.) mentions a third possibility to define degrees. According to his approach, degrees are not equivalence classes resulting

[6] Bierwisch (1987) calls them [-relative].

from the entities compared but degrees are equivalence classes resulting from reference objects such as spatial dimensions. This assumption can also account for the cases where measures from two different scales are compared.

In some approaches gradable adjectives are not analyzed as denoting relations between entities and degrees (type $\langle d, \langle e, t \rangle \rangle$) (e.g., von Stechow, 1984), but as denoting measure functions ($[\![groß]\!] = \lambda$x. μ_{gross}(x)) (e.g., Kennedy, 2007). The more common approach is to analyze gradable adjectives as denoting relations between entities x and degrees d. For this analysis several notional variants exist:

(2.41) a. $[\![groß]\!] = \lambda$d.λx.μ_{gross}(x) $\geq$ d
b. $[\![groß]\!] = \lambda$d.λx.x's height $\geq$ d
c. $[\![groß]\!] = \lambda$d.λx.**groß**(d)(x)
d. $[\![groß]\!] = \lambda$d.λx.x is d-**groß**
e. $[\![groß]\!] = \lambda$d.λx.**groß**(x) = d[7]

Note that there also exist approaches that do not use the notion of degrees (Klein, 1980). These approaches analyze gradable adjectives as denoting partial functions from individuals to truth values. Individuals of which the adjective is true belong to the positive extension of a comparison class, individuals of which the adjective is false belong to the negative extension of the comparison class. The individuals that belong to neither the negative nor the positive extension form the extension gap. Beck et al. (2009) report on languages that do not show evidence for an underlying degree semantics. Motu, spoken in Papua New Guinea, for instance expresses the German comparison constructions in (2.39) in the following way (Beck et al., 2009, ex. 59):

(2.42) Mary na lata, to Frank na kwadoḡi.
Mary TOP tall, but Frank TOP short (TOP = Topic marker)
'Mary is taller than Frank.'

Similar to the German examples in (2.39), the Motu example in (2.42) conveys the information that there is an ordering of the two entities on the height scale such that Mary's position is higher than Frank's position. However, Motu uses the positive form of the adjective to express comparison, thus no degree morphology is present in Motu. Beck et al. (2009) show that there are no other expressions available in Motu that are able to manipulate the degree argument of the adjective indicating that Motu does not have an underlying degree semantics.[8] According to Beck et al. (2009), these crosslinguistic differences show that degrees and scales are a level of abstraction that some languages do not develop such as Motu, and other languages do such as German. Because there are several objections to degree-less approaches in general (see von Stechow, 1984) and

[7] This analysis by Kennedy and McNally (2005) is compatible with analyzing gradable adjectives as being of type $\langle e, d \rangle$.

[8] Other languages for which an underlying degree semantics is not assumed are Washo (Bochnak, 2015), Warlpiri (Bowler, 2016), and Nez Perce (Deal & Hohaus, 2019).

because the focus of this thesis is on German, in what follows I assume an underlying degree semantics for gradable adjectives.

Gradable adjectives such as *klein* ('small') must be interpreted relative to a comparison class as shown in Section 2.2.2. This means they do not denote a property on their own. First, a standard of comparison for the respective comparison class must be determined to evaluate whether an entity is, for example, small or not small. In comparative constructions the degree to which another one is compared provides the standard. In *Dumbo ist größer als Heffalump* ('Dumbo is taller than Heffalump') Heffalump's degree of height serves as the standard of comparison. In the positive unmarked form of a gradable adjective, a covert morpheme (*pos*)[9] fulfills the degree morphology requirement and makes an implicit comparison. This means the *pos*-moprheme takes the adjective denotation and returns a property of individuals (type $\langle e, t\rangle$). I largely follow the anaylsis in Bylinina (2013) but replace the standard degree (d_{ST}) with the function **s** proposed by Kennedy (2007). G is a variable for a gradable degree predicate.

(2.43) $[\![pos]\!] = \lambda \mathrm{G}.\lambda \mathrm{x}.\mathbf{max}\ (\lambda \mathrm{d}.\mathrm{G}(\mathrm{d})(\mathrm{x})) \geq \mathbf{s}(\mathrm{G})$

According to Kennedy (2007, p. 17) **s** is a "contextsensitive function that chooses a standard of comparison in such a way as to ensure that the objects that the positive form is true of 'stand out' in the context of utterance, relative to the kind of measurement that the adjective encodes". The function **s** gives a standard of comparison that takes properties of the gradable adjective and the context of the utterance into account (for a discussion of the advantages of a function **s** over a standard degree see Kennedy (2007)). Because everyone who is for instance tall to a certain degree is also tall to every smaller degree, the *pos*-morpheme must indicate that it is the maximal degree (**max**) that reaches the standard of comparison. The standard of comparison is determined on the basis of the comparison class. The comparison class can be either explicit, i.e., expressed by the noun, or implicit, i.e., provided by the context (von Stechow, 1984). In (2.44a), the noun provides the comparison class, hence the sentence can be paraphrased as in (2.44b).

(2.44) a. Dumbo ist ein kleiner Elefant.
'Dumbo is a small elephant.'

b. Dumbo ist klein für einen Elefanten.
'Dumbo is small for an elephant.'

The comparison class can also be provided by the context when the adjective does not modify a noun as in (2.45a). Thus, the comparison class must be inferred from the context or from world-knowledge. In examples (2.45b) and (2.45c) taken from Partee (2009, p. 14) the adjective modifies a noun phrase, but the denotation of the noun does not serve as the comparison class. Other parts of the sentence (the Agens-NP) influence the meaning of the adjective.

[9] There are several variants of the positive morpheme, which I will not go into here because the choice between them is orthogonal to my purposes.

(2.45) a. Dumbo ist klein.
'Dumbo is small.'

b. Mein 2-jähriger Neffe baute einen großen Schneemann.
'My 2-year-old nephew built a tall snowman.'

c. Die Linguistik-Studierenden bauten einen großen Schneemann.
'The linguistic students built a tall snowman.'

This section showed that subsective/intersective adjectives can be non-gradable or gradable. Non-gradable adjectives denote properties of individuals whereas gradable adjectives denote relations between individuals and degrees. Gradable adjectives have been divided into further subclasses: Bierwisch (1989) distinguishes between dimensional and evaluative adjectives, others distinguish between relative and absolute gradable adjectives (Kennedy & McNally, 2005; Kennedy, 2007). The next two sections explain both classifications. In section 2.3.1 I will relate both classifications and incorporate the entailment properties explained in the previous section.

2.2.3.1 Dimensional and evaluative adjectives

Bierwisch (1989) distinguishes two classes of gradable adjectives: dimensional adjectives (DA) such as *lang, kurz, alt, jung, groß, klein* ('long', 'short', 'old', 'young', 'big', 'small') and evaluative adjectives (EA) such as *faul, fleißig, schön, hässlich* ('lazy', 'industrious', 'pretty', 'ugly'). The example of *begabt* ('skillful') introduced in section 2.2.2 is likely to belong to evaluative adjectives, too. According to Bierwisch (1989), antonymous DAs refer to the same scale of a given dimension but differ in the ordering of the scale; that means both antonyms, e.g., *lang* ('long') and *klein* ('short'), refer to the same scale (LENGTH). To count as long, an entity's length must be greater than or equal to the standard for longness. In contrast, to count as short, its length must be less than or equal to the standard for longness. Crucially, for DAs even the negative antonym expresses a certain degree of length and specifies a positive value of the LENGTH-scale. Unlike DAs, antonyms of EAs refer to different scales or parts of scales. This means that a lazy person for instance need not to have a certain degree of industriousness: a lazy person is not industrious to some extent. Moreover, the antonymy relation is more sytematic for DAs than for EAs. Positive DAs typically have exactly one negative counterpart, whereas for EAs often no unique antonym exists, and sometimes they do not have an antonym at all. Another difference between DAs and EAs lies, according to Bierwisch, in the standard of comparison. Following Cresswell (1976), Bierwisch (1989) argues that the standard of comparison for DAs is the average value of the comparison class[10], hence it is possible to say *Dumbo is smaller than the average of C (e.g. elephants)*. This is less or not appropriate for EAs as in *Hans is more*

[10] Kennedy (2007) illustrates that the calculation of the standard is not as simple as proposed by Cresswell (1976) because the following sentence is felicitous: *Nadia's height is greater than the average height of gymnasts, but she still isn't tall for a gymnast.* Section 2.2.3.2 discusses the standard of comparison in more detail. Chapter 3 presents empirical data regarding the value of the standard.

industrious than the average of C (Bierwisch, 1989, ex. 53c). This difference may also be related to the question of how degrees should be defined for EAs. Unlike for DAs, they cannot be understood as spatial or temporal distances. For EAs, degrees, and hence scales, must be defined in a more general way because they are typically not measurable, i.e., they cannot be mapped straightforwardly on a numerical scale. For instance, Bierwisch (1989) assumes that an evaluative predicate allows different degrees of intensity of the denoted property.

Evaluative adjectives are sometimes claimed to be subjective in nature similar to predicates of personal taste such as *lecker* ('tasty'). For subjective predicates, speakers can disagree on whether the predicate is true of an entity while none of the speakers says something incorrect as demonstrated in the following examples taken from Solt (2018, ex. 1).

(2.46) a. Speaker A: Das Chilli ist lecker! ('The chili is tasty!')
Speaker B: Nein, es is ganz und gar nicht lecker! ('No, it's not tasty at all!')

b. Speaker A: Der Picasso ist schön! ('The Picasso is beautiful!')
Speaker B: Nein, er ist hässlich! ('No, it's ugly!')

For predicates of personal taste, it is typically assumed that an Experiencer or a Judge plays a role for their interpretation (Lasersohn, 2005). However, it has been argued that the subjectivity of evaluative adjectives does not result from an internalized experience which is part of their semantics (Bylinina, 2013). Rather, the evaluative or subjective interpretations are related to the multidimensionality of EAs (Bylinina, 2013; McNally & Stojanovic, 2017; Solt, 2018, 2016; Umbach, 2016). Gradable adjectives can be divided into unidimensional and multidimensional ones. Bierwisch's DAs correspond to unidimensional adjectives. For unidimensional adjectives one criterion (e.g., length) is used to order individuals according to the property the adjective denotes (Klein, 1980; McNally & Stojanovic, 2017). In contrast, for multidimensional adjectives more than one criterion is used to order individuals according to the property the adjective denotes. According to Sassoon (in preparation), many or possibly all of Bierwisch's evaluative adjectives are multidimensional. The multiple dimensions of evaluative adjectives become overt in combination with phrases such as *in jeder Hinsicht* ('in every respect') or *außer/bis auf* ('except for'). This modification is not possible for dimensional adjectives such as *groß* ('big') (Sassoon, in preparation). For every dimension a standard of comparison must be computed and the dimensions must be weighted according to their importance. Thus, subjectivity and potential disagreement arise because speakers can disagree on the respective standards and the relative importance of the single dimensions.

Because both EAs and DAs denote relations between entities and degrees, I assume that the difference between unidimensional and evaluative or multidimensional adjectives is not mirrored in their semantic type. Both adjective classes should be of type $\langle d, \langle e, t \rangle \rangle$, and thus should not differ in their compositional properties. In line with other gradable adjectives such as *groß* ('big'), evaluative adjectives have the following lexical entry:

(2.47) $[\![\textit{schön}]\!] = \lambda \mathrm{d}.\lambda \mathrm{x}.\mu_{\textit{schön}}(\mathrm{x}) \geq \mathrm{d}$

I assume that the differences between DAs and EAs are restricted to their lexical semantic properties. They can differ in the dimensions that have to be included in the interpretation, the nature of the degrees, and whether a Judge/Experiencer is involved.

2.2.3.2 Relative and absolute gradable adjectives

As illustrated in examples (2.39) and (2.40), adjectives are classified as gradable if they can occur in comparative constructions and if they can be modified by degree expressions like *sehr* ('very'), *zu* ('too'), *genug* ('enough'). Differences between gradable adjectives occur with regard to other degree modifiers, with regard to the entailments they license, and with regard to characteristics of vagueness (Kennedy, 2007). These differences result in the distinction between so-called RELATIVE and ABSOLUTE GRADABLE adjectives.

Consider the degree modifiers *fast* ('almost'), *vollständig/komplett* ('completely'), and *ein bisschen* ('slightly').

(2.48) a. Der Boden ist fast/vollständig/??ein bisschen sauber.
'The floor is almost/completely/??slightly clean.'

b. Der Boden ist ??fast/??vollständig/ein bisschen dreckig.
'The floor is ??almost/??completely/slightly dirty.'

c. Der Elefant ist ??fast/??vollständig/??ein bisschen groß.
'The elephant is ??almost/??completely/??slightly big.'

d. Der Elefant ist ??fast/??vollständig/??ein bisschen klein.
'The elephant is ??almost/??completely/??slightly small.'

All four adjectives in (2.48) are gradable because they can be combined with comparative morphology (*sauber-**er**, dreckig-er, größ-**er**, klein-**er***) and they can be modified by other degree modifiers (e.g., *sauber **genug**, **zu** dreckig, groß **genug**, **zu** klein*). However, the degree modifiers in (2.48) are only possible with absolute gradable adjectives. In addition to the differences shown in (2.48), absolute gradable adjectives (examples (2.49) and (2.50)) license different entailments than relative gradable adjectives (examples (2.51) and (2.52)):

(2.49) a. Der Boden ist dreckiger als die Arbeitsplatte. ⇒ Der Boden ist dreckig.
'The floor is dirtier than the countertop.' ⇒ 'The floor is dirty.'

b. Der Boden ist dreckig. ⇒ Der Boden ist nicht sauber.
'The floor is dirty.' ⇒ The floor is not clean.'

(2.50) a. Der Boden ist sauberer als die Arbeitsplatte. ⇒ Die Arbeitsplatte ist nicht sauber.
'The floor is cleaner than the countertop.' ⇒ 'The countertop is not clean.'

b. Der Boden ist sauber. ⇒ Der Boden ist nicht dreckig.
'The floor is clean.' ⇒ 'The floor is not dirty.'

(2.51) a. Dumbo ist größer als Mickey. ⇏ Mickey/Dumbo ist (nicht) groß.
'Dumbo is bigger than Mickey.' ⇏ 'Mickey/Dumbo is (not) big.'

b. Dumbo ist nicht groß. ⇏ Dumbo ist klein.
'Dumbo is not big.' ⇏ 'Dumbo is small.'

(2.52) a. Dumbo ist kleiner als Mickey. ⇏ Mickey/Dumbo ist (nicht) klein.
'Dumbo is smaller than Mickey.' ⇏ 'Mickey/Dumbo is (not) small.'

b. Dumbo ist nicht klein. ⇏ Dumbo ist groß.
'Dumbo is not small.' ⇏ 'Dumbo is big.'

An additional difference between absolute and relative gradable adjectives has been observed with respect to characteristics of vagueness. According to Kennedy (2007), among others, vague adjectives give rise to the Sorites Paradox, vague sentences are context-sensitive, and 'borderline cases' exist for vague sentences. The Sorites Paradox is illustrated in (2.53) and (2.54).

(2.53) P1. Ein Theater mit 1000 Plätzen ist groß. ('A theater with 1000 seats is big.')
P2. Jedes Theater mit einem Platz weniger als ein großes Theater ist groß. ('Any theater with one fewer seat than a big theater is big.')
C. Jedes Theater mit 10 Plätzen ist groß. ('Any theater wit 10 seats is big.')

(2.54) P1. Ein Theater, in dem jeder Platz besetzt ist, ist voll. ('A theater in which every seat is occupied is full.')
P2. Jedes Theater, in dem ein Platz weniger besetzt ist als in einem vollen Theater, ist voll. ('Any theater with one fewer seat than a full theater is full.')
C. Jedes Theater, in dem die Hälfte (oder keiner etc.) der Plätze besetzt ist, ist voll. ('Any theater in which half of (none of etc.) the seats are occupied is full.')

The difference between example (2.53) with a relative gradable adjective and example (2.54) with an absolute gradable adjective is that in (2.53) the two premises (P1 and P2) are true, but the conclusion C is false. The flaw is ascribed to P2, but it is difficult to detect what is wrong with it and why it is accepted (Kennedy, 2007). In contrast, in (2.54) the conclusion is not paradoxical because P2 is already judged to be false.

As illustrated in section 2.2.2, the standard of comparison for gradable adjectives is calculated relative to a comparison class: *Dumbo is small* is interpreted as *Dumbo is small for an elephant*. However, the value of the comparison class is not relevant for all gradable adjectives. It is only relevant for relative gradable adjectives such as *groß* and *klein*. Thus, these adjectives are context-sensitive. For absolute gradable adjectives such as *sauber* ('clean'), *dreckig* ('dirty') or *voll* ('voll') the value of the comparison class is not relevant. There are two variants of absolute gradable adjectives: minimum standard and maximum

standard absolute gradable adjectives. The arguments of minimum standard adjectives such as *dreckig* ('dirty') must exhibit the property denoted by the adjective to a minimal/non-zero degree. The arguments of maximum standard adjectives such as *sauber* ('clean') must exhibit the property denoted by the adjective to a maximal degree. This is indicated by the contradictory examples (2.55a) and (2.55b).

(2.55) a. # Der Boden ist nicht dreckig, jedoch ist dort ein Fleck.
'The floor is not dirty, though there is a spot.'

b. # Der Boden ist SAUBER, aber er könnte noch sauberer sein.[11]
'The floor is CLEAN, but you can make it cleaner.'

Dreckig ('dirty') in (2.55a) is a minimum standard absolute gradable adjective, hence the standard of comparison is a minimal degree of dirtiness. As a consequence, a denial as in the first part of the sentence in (2.55a) should entail that the floor possesses no amount of dirtiness at all. Therefore, the utterance in (2.55a) is contradictory (Kennedy, 2007). *Sauber* ('clean') in (2.55b) is a maximum standard absolute gradable adjective, hence the standard of comparison is a maximal degree of cleanliness. Thus, the first part of the sentence in (2.55b) should entail that the floor has a maximum amount of cleanliness, i.e., it cannot be cleaner than it is. Therefore, the sentence in (2.55b) is contradictory (Kennedy, 2007).

The observation that utterances with relative gradable adjectives like (2.56a) and (2.56b) are not contradictory points to a standard degree around the center of the scale.

(2.56) a. Dumbo ist nicht groß, aber seine Größe ist normal für sein Alter.
'Dumbo is not tall, but his heigth is normal for his age.'

b. Mickey ist klein, aber es gibt kleinere Mäuse.
'Mickey is small, but there are smaller mice.'

A midpoint standard for relative gradable adjectives explains why negation of a property allows a positive value of the property as in (2.56a) in contrast to minimum standard adjectives. And it accounts for why higher values of the property are possible as in (2.56b) in contrast to maximum standard adjectives (Kennedy, 2007).

Morzycki (2016) states that evaluative adjectives have minimal standards because they license the same inferences as shown in (2.49a).

(2.57) Clyde ist dümmer als Floyd. ⇒ Floyd ist dumm.
'Clyde is more stupid than Floyd.' ⇒ 'Floyd is stupid.'

However, a comparison of this example with (2.49a) reveals that they are slightly different. The inference in (2.49a) says something about the entity which is compared to another entity. The inference in (2.57) says something about the entity to which another

[11] The capitals signal focal stress. According to Kennedy (2007), this forces a precise interpretation of the maximal gradable adjective, which is necessary for this kind of entailment. For a discussion on imprecise meanings of maximal absolute gradable adjectives see Kennedy (2007).

one is compared. Bierwisch (1989) already noticed that evaluative adjectives differ in the entailments they license. For some of them, it is possible to infer that the property denoted by the adjective holds of both entities, for other adjectives this is less clear. Sassoon (2013) provides a typology of multidimensional adjectives and shows that some of them have absolute standards, whereas others have relative (or midpoint) standards like relative gradable adjectives. As shown in the previous section, multidimensional adjectives belong to the class of evaluative adjectives. Hence, I follow Sassoon (2013) and assume that the distinction into absolute and relative gradable can be applied to both dimensional and evaluative adjectives.

The third characteristic of vague predicates is the existence of 'borderline cases'. Borderline cases exist for relative gradable, but not for absolute gradable adjectives. *Groß* ('big') is a case in point: in any given context there is a set of objects that can be clearly judged as big, and another set of objects that can be judged as not big. In addition, there are also objects that are less well-defined. Thus, in antonym pairs the negation of one adjective does not entail the assertion of the other adjective as displayed in the above examples (2.51b) and (2.52b). In contrast, such inferences are licensed for absolute gradable adjectives (see the above examples (2.49b) and (2.50b)). This is because for absolute gradable adjectives the minimal positive degree corresponds to the maximal negative degree on the same scale (Kennedy, 2007). The infelicitous inferences in (2.51b) and (2.52b) result from the context-sensitive standards of relative gradable adjectives (Klein, 1980): the standard for *groß* ('big') and *klein* ('small') need not to be the same degree (Kennedy, 2007). This is also pointed out by Solt (2011). If the negative antonym is analyzed as the direct oppsosite of the positive antonym, then the standard for both would be the same single degree, that completely divides the scale into big and small entities. However, there are sizes that are judged as neither big nor as small, as suggested by the felicity of sentences like *Dumbo ist nicht groß, aber er ist auch nicht klein* ('Dumbo isn't tall, but he's not short either'). As a consequence, one would have to assume two different standards for *groß* ('big') and *klein* ('small') with a gap in between. According to Solt (2011), there should be nothing that rules out the possibility that the position of the two standards can be reversed. In other words, that the standard of the negative antonym is at a higher position than the standard of the positive antonym. As a result, sentences like *Dumbo ist sowohl groß als auch klein für einen Elefanten* ('Dumbo is both tall and short for an elephant') should be true in some contexts. However, the standards for *groß* ('big') and *klein* ('small') stand always in the same relation, such that the standard for *groß* ('big') should never be below the standard for *klein* ('small') with respect to the same entity. To account for these observations, Solt (2011) defines the standard for relative gradable adjectives as a range of degrees rather than a single degree. Solt (2011) notes that her analysis of the standard is restricted to relative gradable adjectives. Because borderline cases do not exist for absolute gradable adjectives, the standard should be a single degree on the scale, which is shared by the positive and the negative antonym. This has also been claimed by Rotstein and Winter (2004), although they illustrate that in some contexts minimum and maximum standard

adjectives[12] can be interpreted as non-complementary. For a moist towel for example it is difficult to tell whether it is wet or dry. The same holds for the antonym pair *clean/dirty*: in example (2.58) (taken from Rotstein & Winter, 2004, ex. 10b) there seems to be a degree of dirtiness for which both *clean* and *dirty* would be judged false. Nevertheless, the default is to interpret antonym pairs as complementaries.

(2.58) This glass is almost dirty: It is certainly not clean, since it has some small spots on it, but it is not really dirty, and I am willing to drink from it if you insist.

In summary, gradable adjectives can be classified as relative and absolute gradable adjectives. Relative gradable adjectives show properties of vague predicates. Due to their context-sensitivity, their standard of comparison is assumed to be located around the center of the scale. Antonym pairs of relative gradable adjectives are 'contraries' rather than 'complementaries' (see Solt, 2011), i.e., there is for instance a range of sizes which count as neither big nor small. Absolute gradable adjectives on the other hand are not vague. It has been argued that their standard of comparison is located at an endpoint of the scale. In addition, it has been argued that they are 'complementaries', i.e., the standard of comparison is a single degree, which is shared by the positive and the negative antonym.

Despite the differences, relative and absolute gradable adjectives behave alike with respect to their acceptability in comparatives and with respect to some degree modifiers. Therefore, relative and absolute gradable adjectives should have the same semantic type ($\langle d, \langle e, t \rangle \rangle$), i.e., they should both denote relations between entities and degrees (see Kennedy & McNally, 2005; Kennedy, 2007). Various sources for the differences between relative and absolute gradable adjectives have been discussed in the literature. In the following, I briefly describe three different approaches (Kennedy, 2007; Kennedy & McNally, 2005; McNally, 2011; Toledo & Sassoon, 2011). As far as I can see, they all make the same predictions for the interpretation of relative and absolute gradable adjectives.

Kennedy and McNally (2005) and Kennedy (2007) ascribe the differences between gradable adjectives to differences in their scalar structure. In their analysis, gradable adjectives belong to one of the four scale types in Figure 2.1.

[12] In Rotstein and Winter (2004) minimum standard adjectives are called 'partial' adjectives, maximum standard adjectives are called 'total' adjectives.

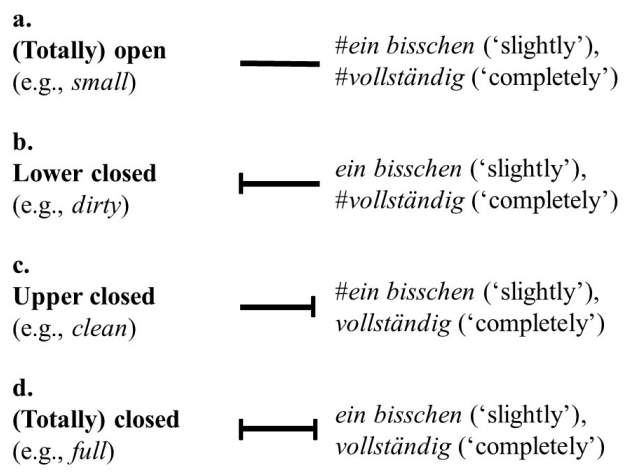


Figure 2.1: Typology of scale structure (from Kennedy, 2007, p. 33; Toledo and Sassoon, 2011, p. 136).

The scale type can account for the distribution of degree modifiers such as *ein bisschen* ('slightly') and *vollständig* ('completely') illustrated in (2.48) above. *Ein bisschen* ('slightly') takes the minimum degree of the scale, hence it is only applicable if the scale has a minimal endpoint as in b. and d. in Figure 2.1. In contrast, *vollständig* ('completely') takes the maximum degree of the scale, thus it is only applicable if the scale has a maximal endpoint as in c. and d. in Figure 2.1. Open scales as in a. in Figure 2.1 have no endpoints at all, therefore they cannot be modified by neither *ein bisschen* ('slightly') nor *vollständig* ('completely'). Kennedy and McNally (2005) state that the standard of comparison is context-dependent if the adjective has an open scale. If the adjective has a (partially) closed scale, one of the endpoints serves as the standard of comparison. To explain how the standard is fixed appropriately, Kennedy (2007, p. 36) proposes an economy principle (Interpretive Economy) that states: "Maximize the contribution of the conventional meanings of the elements of a sentence to the computation of its truth conditions." This means the scale structure, which is part of the adjective's conventional meaning, takes priority over contextual factors in the determination of the standard. As a consequence, a context-dependent standard is only calculated if the adjective's scale lacks an endpoint.

This account raises the question of whether the scale type is represented in the denotation of the adjective. If this were the case, the *pos*-morpheme must be defined differently for each scale type to ensure that the appropriate standard is determined. However, this assumption stands in conflict with the suggestion that all gradable adjectives have the same semantics. By defining the *pos*-morpheme as in (2.43), the function **s** satisfies the scale type by definition. As a consequence, the denotation of the adjective itself does not contain all meaning components of the adjective.

McNally (2011) objects that it is not possible to define a unified function **s**. Therefore, she claims that the role of the scale type plays a too important role in the other approaches. She observes two problems with Kennedy's (2007) analysis. First, there are adjectives

that are interpreted with non-endpoint standards although they have closed scales (e.g., *familiar*). Second, some adjectives have absolute standards, but the standard is not located at the endpoint of the scale (e.g., *full*). McNally (2011) brings the (alleged) primary function of adjectives into play: adjectives sort individuals according to the way they exhibit a given property. McNally (2011) suggests that the way individuals manifesting a given property to different degrees are grouped, determines the standard for the adjective that refers to the property. She claims that individuals can be classified by similarity or by rule. In her analysis, relative gradable adjectives classify individuals by similarity, whereas absolute gradable adjectives classify individuals by rule. Classification by similarity requires only a partial match between the classification criteria and the individual's properties. Classification by similarity means that a specific individual or a property of that individual is compared to the representation of another equally specific individual or properties of that individual. Based on examplars that have been previously identified as big or small, it is possible to classify any other individual as big or small relative to the known exemplars. Hence, the comparison class is important because it provides the exemplars on which the classification is based.

In contrast, classification by rule requires a strict matching between the classification criteria and the individual's properties. Classification by rule means that a representation of a specific individual (e.g., the degree of fullness of a wine glass) is compared to a more abstract representation (e.g., the degree of fullness for glasses in general). Because a simple rule applies, i.e., the property holds or it does not hold, no comparison to other individuals is necessary. Therefore, no comparison class is needed. McNally (2011) concludes that abstract scale properties are only indirectly related to the differences between relative and absolute gradable adjectives. For her, the crucial difference is whether it is possible to establish clear applicability conditions for the rule. According to her, an important factor is the ease with which the degree that constitutes the standard can be perceived, e.g., the minimum or maximum degree of the scale.

Toledo and Sassoon (2011) argue that both relative and absolute gradable adjectives are interpreted relative to a comparison class and are subject to contextual considerations. This account can explain for example the intuition that *clean* is understood differently in the context of kitchen knives *versus* in the context of surgical instruments, or why *full* is understood differently in the context of wine classes compared to oil tanks. According to Toledo and Sassoon (2011), the difference between relative and absolute gradable adjectives results from the nature of the comparison class. For relative gradable adjectives, the comparison class consists of no two different counterparts of the same individual such as possible temporal stages. It comprises other members of the same category as the individual. Whether the property holds of an individual, is evaluated by comparing that individual to other distinct individuals ('between-individual interpretation'). This procedure resembles McNally's (2011) classification by similarity. For absolute gradable adjectives, the comparison class includes only different counterparts (possible temporal stages) of the same individual, hence only one individual contributes values to the comparison. This is

what Toledo and Sassoon (2011) call 'within-individual interpretation': for instance, the description of a shirt as dirty is based on the visualization of this shirt in various degrees of dirtiness rather than comparing the degree of dirtiness of the shirt with the degree of dirtiness of other shirts. The standard of comparison is determined via the comparison class, an economy principle similar to Kennedy's (2007) and a grammaticalization principle. The grammaticalization principle adds that the type of standard that is usually selected for an adjective is encoded as the default and applied whenever possible. Toledo and Sassoon (2011) suggest that the correct type of comparison class is selected by taking other properties of adjectives into account. Absolute gradable adjectives are associated with stage-level predicates. They denote transient values that equal the variance within individuals. Relative gradable adjectives are associated with individual-level predicates. They denote enduring values that encode no sufficient variance within individuals.

I claim that all three approaches are similar in that the differences between relative and absolute gradable adjectives result from lexical semantic properties. Lexical semantic properties can be defined as scalar structure (Kennedy & McNally, 2005; Kennedy, 2007), classification procedures (McNally, 2011) or the nature of the comparison class (Toledo & Sassoon, 2011). As a consequence, the denotation of a gradable adjective does not contain all meaning components, but additional information is necessary to interpret it correctly.

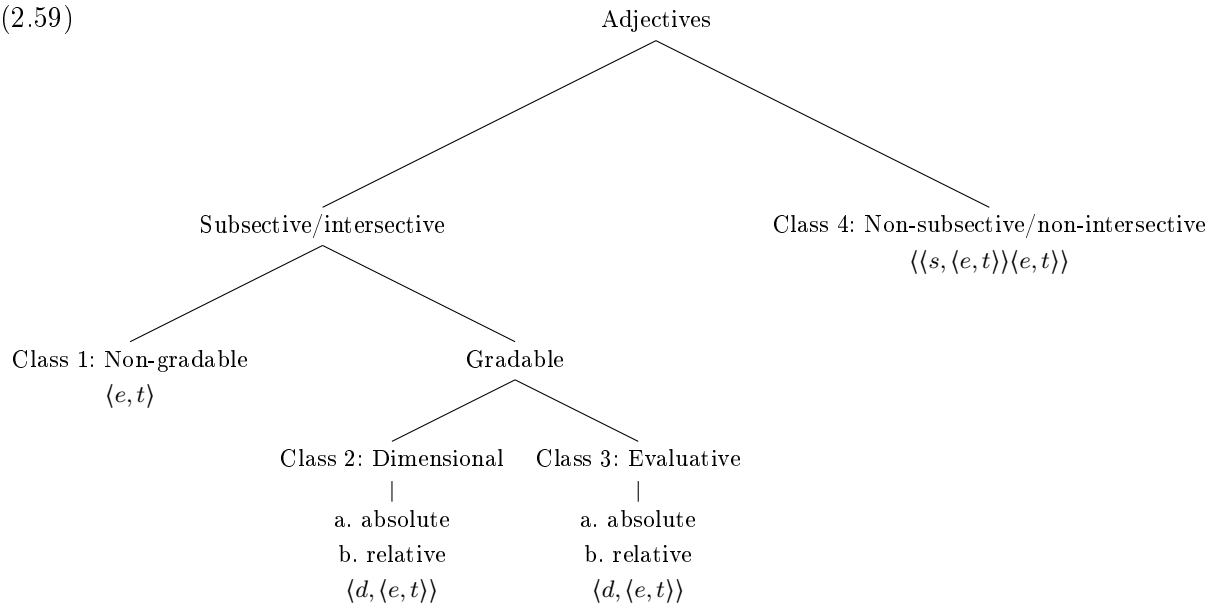
In Section 2.3, I propose a classification of adjectives that extends the classification based on the entailment properties in (2.38) by the different gradability properties introduced in this section. Moreover, I argue that the resulting adjective classes differ in their semantic complexity.

2.3 Semantic properties and complexity

2.3.1 Entailment and gradability properties

An important classification of adjectives functioning as nominal modifiers is according to the inferences they license. In Section 2.2.2, I concluded that two types of adjectives exist: subsective/intersective adjectives (e.g., *männlich, groß, begabt*) and non-subsective/non-intersective adjectives (e.g., *ehemalig, angeblich*). Section 2.2.2 also illustrated that differences among the subsective/intersective adjectives exist. In Section 2.2.3, I suggested that the differences among the adjectives summarized as subsective/intersective can be captured by their gradability properties. Therefore, it is crucial to incorporate entailment and gradability properties to notice the subtle differences between adjective classes. Taking subsectivity, intersectivity, and gradability properties (dimensional, evaluative, relative gradable, and absolute gradable) into account results in 36 logical possible combinations. Given that some properties depend on each other, e.g., only gradable adjectives are dimensional or evaluative, only the following six classes are attested.

(2.59)



As mentioned in Section 2.2.3, I use gradability as a property to further distinguish subsective/intersective adjectives. Recall that all non-subsective/non-intersective adjectives are non-gradable, therefore gradability is not applied to this class in (2.59). The semantic literature on adjectives provides different diagnostic criteria for the assignement of adjectives to one of the classes in (2.59) (e.g., Bierwisch, 1989; Kamp & Partee, 1995; Kennedy & McNally, 2005; Sassoon, in preparation). D1 in Figure 2.2 differentiates between subsective/intersective and non-subsective/non-intersective adjectives. D2 distinguishes between non-gradable and gradable subsective/intersective adjectives. D3 divides gradable adjectives into dimensional and evaluative adjectives. This diagnostic criteria is based on the assumption that evaluative adjectives are often multidimensional and can be modified by phrases such as *in jeder Hinsicht* ('in every respect') or *außer/bis auf* ('except for') (Sassoon, in preparation). D4 discriminates between absolute and relative gradable adjectives. As shown in section 2.2.3.2, absolute but not relative gradable adjectives can be modified by *vollständig* ('completely') or *ein bisschen* ('slightly').

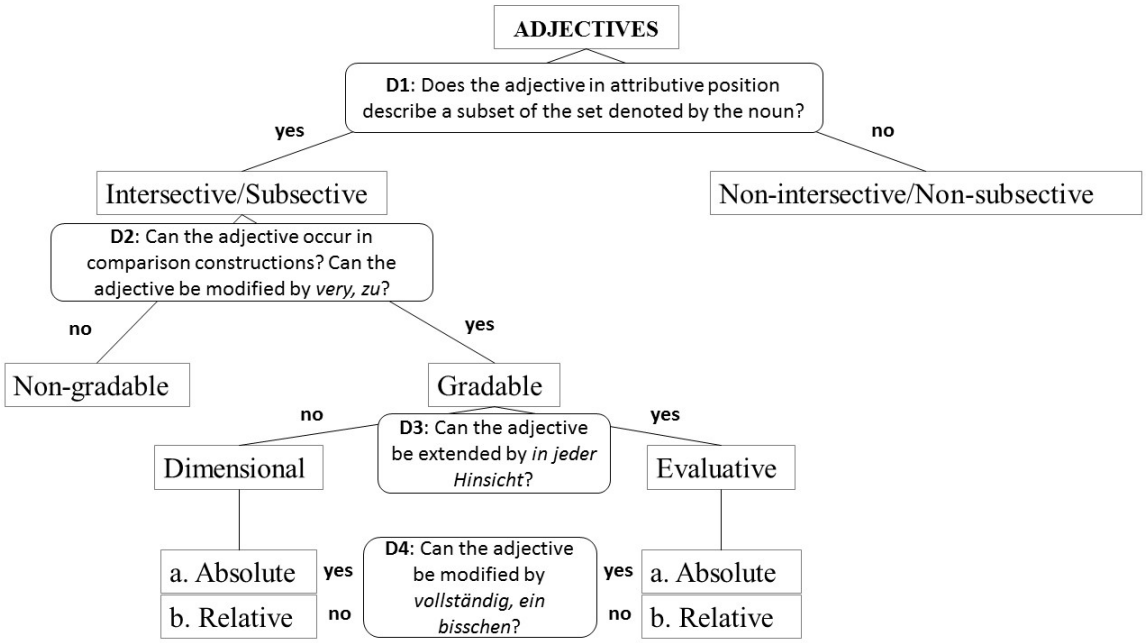
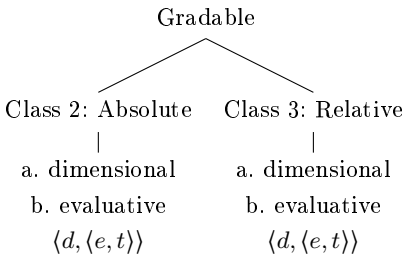


Figure 2.2: Classification criteria D1-D4 for adjectives.

Note that in principle the node of gradable adjectives in (2.59) could be branched differently such that for instance absolute gradable adjectives constitute Class 2, which consists of (a) dimensional and (b) evaluative adjectives, and relative gradable adjectives constitute Class 3, which also consists of (a) dimensional and (b) evaluative adjectives as in (2.60).

(2.60)



I assume the classification in (2.59) for reasons that relate to the notion of semantic complexity, which is explained in the following. The semantic types in the classification proposed in (2.59) reveal some differences in complexity between the classes. However, complexity is not always restricted to compositional semantics. That is, some adjective classes seem more complex than others, but this difference is not mirrored in the semantic type as mentioned in section 2.2.2 (see Morzycki, 2016). Assuming that there are differences in complexity between all the classes proposed in (2.59), or at least between the four main classes, the question is what the source of this differences is, in other words, how complexity is defined. The general issue regarding this question is that there is no independent measure of complexity and there is no agreement on how to properly define complexity (Culicover, 2013). Inferences about what is more or less complex are mostly based on data from processing, acquisition, language change, and variation. Culicover (2014) argues for

a distinction between formal and processing complexity: formal complexity refers to the idiosyncrasy and irregularity involved in a construction. Processing complexity refers to the resources required by a speaker to map form and meaning. Menn and Duffield (2014) point out that these two types of complexity need not be identical. Compared to morphological or syntactic complexity, semantic complexity has been discussed even more rarely (but see Matthewson, 2014). Moreover, research dealing with complexity investigates whether languages differ in complexity by comparing typologically different languages. Matthewson (2014) compares St'át'imcets, a Northern Interior Salish language spoken in the southwest interior of British Columbia, and English with respect to quantification. I do not give the details of her analysis of the two languages, but focus on her approach to semantic complexity. Matthewson (2014) remarks that there are different levels at which semantic complexity can be measured based on what a speaker must know of a language. This is (i) the lexical entries for all morphemes in the language, hence complexity can be measured with respect to the formal representation of lexical entries. In addition, a speaker must know (ii) the compositional mechanisms of a language, thus complexity can refer to the calculation of truth conditions for utterances. And (iii), speakers must be able to interpret pragmatic inferences or context-dependent meaning. As an additional measurement, Matthewson includes type-complexity. Matthewson exemplifies three metrics: length of description, semantic strength, and type complexity.[13]

For length of description, Matthewson (2014) uses an example by Gennari and Poeppel (2003). They claim that verbs denoting externally caused events (e.g., *break*) are more complex than verbs denoting internally caused events (e.g., *grow*) because the former require two participants. Therefore, they have longer lexical entries. According to Matthewson (2014), the problem with this metric is that often there are many different ways to write logically equivalent semantic formulas that differ in length, but express the same meaning. With respect to adjectives, alternative semantic formulas that express the same meaning exist. However, if the length of the λ-prefix is considered, the semantic formulas within a class are consistent regarding their length. Hence, I assume that the problem noted by Matthewson (2014) can be neglected for the present purpose. Regarding semantic strength, Gennari and Poeppel (2003) define that elements are more complex if they have more entailed properties than simpler elements. Matthewson (2014) criticizes that in some cases this metric can be reduced to length of description. In other cases, e.g., numerals, one element entails another element (*four* entails *three*, *three* entails *two* etc.), but they do not differ in complexity. Regarding type complexity, Matthewson (2014) states that in English quantified noun phrases are of type $\langle\langle e,t\rangle,t\rangle$ (generalized quantifiers), hence determiners must be of type $\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$, they take two set arguments. In contrast, in St'át'imcets quantifiers are of type $\langle\langle e,t\rangle,\langle e,t\rangle\rangle$, they take one set argument and return a smaller set. Therefore, quantification in St'át'imcets is less complex than quantification in English ac-

[13] Matthewson mentions other metrics like paradigm complexity and expressing complexity. These two metrics are more relevant for the comparison of complexity across languages. Since the present purpose is to identify differences within a specific domain of a single language, I forgo these measurements.

cording to Matthewson (2014). Note that it is not explained how exactly type complexity is calculated.

The three metrics length of description, semantic strength, and type complexity are all related to the formal representations of lexical entries. Matthewson (2014, footnote 10) remarks that she sets complexity of composition mechanisms aside. Regarding adjectives, one would have to define whether predicate modification is less complex than functional application. I am not able to answer this question for now and therefore follow Matthewson (2014) and do not take it into account. However, it may be possible that in this regard semantic theory can benefit from acquisition data if one mechanism turns out to be the default.

With regard to pragmatic inferences, Matthewson assumes for conversational implicatures that they do not add formal complexity since they are not represented in the truth conditions. However, intuitively an utterance with an implicature should be judged as more complex than an utterance without an implicature. Adjectives do not involve pragmatic inferences, but sometimes context-dependent meaning. I propose that the context-dependent meaning, and other meaning-contributing components are not formally implemented, but have to be stored separately in the lexicon.

Adjectives affect all three aspects mentioned by Matthewson (2014) speakers must know of a language: lexical entries, compositional mechanisms, and partially context-dependent meaning. On the basis of the differences between the adjective classes that have been proposed in the semantic literature and that are mirrored in the classification in (2.59), I propose that semantic complexity is a combination of the following three factors:

(2.61) Semantic complexity

(i) length of description: number of variables

(ii) type complexity: type level

(iii) lexical properties which contribute to the adjective's meaning

Type complexity is defined by means of the type's level. I assume that the set of types T includes: 1) atomic types e (individuals), t (truth-values), d (degrees), s (worlds) $\in T$; 2) functional types such that if $\sigma, \tau \in T$, then $\langle \sigma, \tau \rangle \in T$. The level $\#(\sigma)$ of the corresponding type σ is defined as follows (Ede Zimmermann, p.c.):

(2.62) Type level

(a) $\#(e) = \#(t) = \#(d) = \#(s) = 0$

(b) $\#(\langle \sigma, \tau \rangle) = \max(\#(\sigma) + 1, \#(\tau))$

In general, an adjective is more complex than another adjective if its denotation includes more variables, if its type is of a higher level, and if additional lexical information is required to access the adjective's complete meaning. This definition of complexity fits well for the classification proposed in (2.59). To be more precise, some type τ is more complex than a type σ iff τ has a higher type level than σ, or iff σ and τ have the same type level but τ

has a higher number of variables than σ. This results in the lexicographic order of ranks $\rho(\sigma) = (\#(\sigma)$, number of variables$(\sigma))$ in (2.63) (Ede Zimmermann, p.c.).

(2.63) $\rho(\langle e,t\rangle) = (1,1) < \rho(\langle d,\langle e,t\rangle\rangle) = (1,2) < \rho(\langle\langle s,\langle e,t\rangle\rangle\langle e,t\rangle\rangle) = (2,2)$

This order does not include the third metric, namely the lexical properties that contribute to the adjective's meaning. It is open how exactly lexical properties interact with compositional properties and whether the properties are of different relevance for semantic complexity. In Table 2.1, I describe the adjective classes from (2.59) regarding all three metrics. I claim that Class 1 is the least complex class and Class 4 is the most complex class. The subclasses a. and b. are equally complex. This results in the *Semantic Complexity Hierarchy* in Table 2.1. In the remainder of the thesis I sometimes name the classes like one of the adjectives in the *Examples* column in Table 2.1 for ease of understanding. Note that each adjective class is assumed to comprise all adjectives that share the same specification of properties.

If the alternative classification in (2.60) is assumed, one would have to explain why absolute gradable adjectives should be more complex than relative gradable adjectives or *vice versa*, and why dimensional and evaluative adjectives should be equally complex. However, neither assumption follows from the definition of complexity in (2.61).

2.3.2 Notional properties

In Section 2.3.1, I proposed a classification of adjectives that incorporates entailment and gradability properties. Recall that in Section 2.2.1 I described an additional classification of adjectives according to notional properties. How are these properties related? According to Demonte (2011), notionally-based classes need not to be identical with the classes emerging from formal semantic considerations such as entailment and gradability properties. It may be that notional classes naturally follow from formal semantic properties (McNally, 2016). However, it is unclear how notional and semantic classifications correspond because there is only little research on the connection between these classification systems (see Morzycki, 2016). Taking adjectives from the notional classes proposed by Dixon (1982) Dimension, Physical Property, Color, Human Propensity, Age, Value, and Speed it seems that they can be transferred into the semantic classes in (2.59), but that the correspondence between the classifcations may not be a strict one. First of all, all notional classes are subsective/intersective. Assuming that color words are non-gradable they can be clearly assigned to Class 1. Adjectives from the notional classes Dimension (e.g., *tall*), Age (e.g., *old*), and Speed (e.g., *fast*) are relative gradable dimensional adjectives and hence belong to Class 2b. Classes that are more heterogeneous like Physical Property fall into various semantic classes. For instance, *triangular* is non-gradable and belongs to Class 1, *dirty* is absolute gradable dimensional and belongs to class 2a, *soft* is relative gradable dimensional and belongs to Class 2b, and *beautiful* is relative gradable

Table 2.1: *Semantic Complexity Hierarchy* based on compositional and lexical semantic properties of adjective classes 1 to 4.

Adjective class	Number of variables	Semantic type	Lexical properties	Examples
Class 1 'BLUE-class' subsective/intersective, non-gradable	1 (individual variable)	$\langle e, t\rangle$ (Level-1)	/	*blau* ('blue'), *deutsch* ('German'), *männlich* ('male'), *schwanger* ('pregnant')
Class 2a 'CLEAN-class' subsective/intersective, absolute gradable, dimensional	2 (individual and degree variable)	$\langle d, \langle e, t\rangle\rangle$ (Level-1)	Scalar structure	*sauber* ('clean'), *dreckig* ('dirty'), *trocken* ('dry'), *nass* ('wet'), *voll* ('full'), *leer* ('empty')
Class 2b 'BIG-class' subsective/intersective, realtive gradable, dimensional	2 (individual and degree variable)	$\langle d, \langle e, t\rangle\rangle$ (Level-1)	Scalar structure	*groß* ('big'), *klein* ('small'), *lang* ('long'), *kurz* ('short'), *teuer* 'expensive', *billig* ('cheap')
Class 3a 'HEALTHY-class' subsective/intersective, absolute gradable, evaluative	2 (individual and degree variable)	$\langle d, \langle e, t\rangle\rangle$ (Level-1)	Scalar structure, abstract scales, multiple dimensions, judge/experiencer	*gesund* ('healthy'), *krank* ('sick'), *gefährlich* ('dangerous'), *sicher* ('safe'), *langweilig* ('boring')
Class 3b 'BEAUTIFUL-class' subsective/intersective, relative gradable, evaluative	2 (individual and degree variable)	$\langle d, \langle e, t\rangle\rangle$ (Level-1)	Scalar structure, abstract scales, multiple dimensions, judge/experiencer	*schön* ('beautiful'), *hässlich* ('ugly'), *intelligent* ('intelligent'), *fleißig* ('industrious')
Class 4 'FORMER-class' non-subsective/non-intersective (non-gradable)	2 (individual and situation/ world variable)	$\langle\langle s, \langle e, t\rangle\rangle\langle e, t\rangle\rangle$ (Level-2)	?[a]	*ehemalig* ('former'), *zukünftig* ('future'), *angeblich* ('alledged'), *möglich* ('possible')

[a] For class 4, it is open whether lexical properties must be defined independent of the adjectives' lexical entries and how they look like. One possibility could be the fact that these adjectives are not subsective, i.e., they do not create a subset of the noun denotation.

evaluative and belongs to Class 3b. These examples suggest a relation between notional and semantic classes, but to conclude that notional classes naturally follow from formal semantic properties (McNally, 2016) may be to strong. It seems plausible to assume that notional classes express the concepts that compose the mental lexicon. Then, it would be expected that these concepts need not necessarily expressed only by adjectives. Semantic classes describe the compositional and lexical meaning of these concepts, i.e, properties that guide compositional operations and interpretive features, for instance whether an adjective must be interpreted relative to a standard of comparison. In Chapter 4, I present empirical data concerning the relation between notional and semantic classes.

2.4 Implications for the acquisition of the semantics of adjectives

The classification of adjectives proposed in Section 2.3.1 takes entailment and gradability properties into account. Regarding the acquisition of adjectives, this classification raises the question in (Q):

(Q) How do monolingual German-speaking children acquire the semantics of adjectives?

The aim of this thesis is to postulate an acquisition path for the semantics of adjectives. Based on the assumption that the adjective classes proposed in Section 2.3.1 differ in semantic complexity, I hypothesize that children acquire the semantics of adjectives in an order determined by semantic complexity as stated in (H1).

(H1) The order of acquisition of adjectives is determined by their semantic complexity.

In the following, I motivate this hypothesis and present the research questions that are addressed in the empirical part of this thesis to test H1. Complexity in various manifestations has been considered as a predictor for the order of acquisition of syntactic and phonological phenomena. Hawkins (1987) formulates his predictions for the order of acquisition based on implicational universals of the form "If a language has some property P, then it has also some other property Q". Implicational universals typically capture cross-linguistic variation, but according to Hawkins (1987) they can also make predictions for language acquisition. With regard to the order of acquisition, Hawkins (1987) argues that the acquisition of the consequent property of an implicational statement (Q) will either precede or occur simultaneously with the acquisition of the antecedent property of the implicational statement (P), i.e., the property that a language has in addition to another property. The acquisition of property P before property Q is ruled out. Hawkin's second prediction states that in stages in which both properties are acquired property Q should have a quantitative advantage over property P. In other words, either property Q will be more successfully produced and comprehended, or property Q and property P will be equally succesfully

produced and comprehended. The more succesful production and comprehension of property P over property Q is ruled out. This second prediction can for instance be tested by the number of errors children make. Mainly on the basis of data from the acquisition of different phonemes, Hawkins (1987) concludes that the order-of-acquisition prediction and the quantitative prediction are borne out. Hawkins (1987) also remarks that the quantitative prediction logically implies the order-of-acquisition prediction, i.e., it is not possible that the quantitative prediction is true while the order-of-acquisition prediction is false. In contrast, the order-of-acquisition prediction does not imply the quantitative prediction (see Hawkins, 1987).

Unlike Hawkins (1987), Snyder (2007) does not derive his predictions for language acquisition from cross-linguistic comparison but focuses on single languages. His predictions for the setting of mainly syntactic parameters are the following (Snyder, 2007, p. 7):

1. "If the grammatical knowledge (including parameter settings and lexical information) required for construction A, in a given language, is identical to the knowledge required for construction B, then any child learning the language is predicted to acquire A and B at the same time."

2. "If the grammatical knowledge (including parameter settings and lexical information) required for construction A, in a given language, is a proper subset of the knowledge required for construction B, then the age of acquisition for A should always be less than or equal to the age of acquisition for B. (No child should acquire B significantly earlier than A.)"

Snyder (2007, p. 8) further assumes that children's spontaneous speech is characterized by grammatical conservatism, i.e., children do not make productive use of a new grammatical construction before they have not figured out that the construction exists in adult language and before they have not identified the adults' grammatical basis for it. As a consequence, a rapid transition from no use of a construction to a frequent and correct use of this construction can often be observed. Hawkins' (1987) and Snyder's (2007) accounts assume that the order of acquisition in phonology and syntax is influenced by the parameter settings either across languages or within one language. Both accounts imply that additional material, e.g., a property or knowledge, is not acquired before the property or knowledege this additional material builds on. Although not explicitly mentioned, it can be inferred that this additional material increases the complexity of the respective construction.[14]

[14] Other accounts define syntactic complexity more concretely. In an early proposal by Brown and Hanlon (1970), a sentence B has a greater cumulative derivational complexity than a sentence A if the derivation of Y follows all the rules applied in the derivation of X plus at least one rule not applied to X. Brown and Hanlon (1970) predict that less complex structures are acquired before more complex one. The authors claim that the prediction is borne out with respect to the emergence of different sentence structures in the spontaneous speech of three children. Frank (1998) assumes that syntactic operations can differ with respect to complexity. It is expected that structures requiring more complex operations are acquired later than structures requiring less complex operations, because initially more complex operations are not present in children's grammar.

As can be seen from these two approaches, complexity has been considered as determining the order of acquisition in the area of syntax and phonology. To date, semantic complexity has received little attention in acquisition research. Studies taking semantic properties into account either lack a clear definition of semantic complexity from which hypotheses for acquisition can be derived (Blackwell, 2005; Tribushinina et al., 2014) or are restricted to particular adjective classes such as gradable adjectives (Hohaus, Tiemann, & Beck, 2014). Blackwell (2005) and Tribushinina et al. (2014) investigated the emergence of notional adjective classes in spontaneous speech. According to Blackwell (2005), adjectives with a noun-dependent meaning are acquired later than adjectives with a noun-independent meaning. Tribushinina et al. (2014) claim that adjectives with so-called 'concrete semantics', i.e., referring to prominent concepts with perceptually grounded meanings, emerge earlier than adjectives referring to abstract properties, which are argued to be less accessible to the child. However, both explanations remain rather vague and are, as noted by Blackwell (2005) herself, only tentative. Hohaus et al. (2014) predict the order of acquisition for gradable adjectives and comparison constructions based on parametric variation.

The classification of adjectives and the definition of semantic complexity I proposed in Section 2.3.1 allow predictions for the acquisition of different adjective classes. In line with Snyder (2007), hypothesis H1 predicts adjectives to be acquired in the following order: less complex adjectives in the *Semantic Complexity Hierarchy* should be acquired before or simultaneously with more complex adjectives in the *Semantic Complexity Hierarchy*. Equally complex adjectives should be acquired at the same age. In line with Hawkins (1987), I assume that it is possible that semantic complexity influences the order of acquisition whithout affecting the frequency of succesful adjective use or comprehension.

For the production of adjectives in spontaneous speech the following research questions are addressed in Chapter 4:

(P1) Is the *Semantic Complexity Hierarchy* reflected in the mean age of acquisition for the adjective classes?

(P2) Is the *Semantic Complexity Hierarchy* reflected in the growth patterns of the adjective classes?

If the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition (P1), Class 1 (BLUE-class) should have a lower or the same mean age of acquisition than Class 2. Class 2 should have a lower or the same mean age of acquisition than Class 3, and Class 3 should have a lower or the same mean age of acquisition than Class 4 (FORMER-class). Classes 2a (CLEAN-class) and 2b (BIG-class) should not differ in their mean age of acquisition because they are equally complex. Classes 3a (HEALTHY-class) and 3b (BEAUTIFUL-class) should not differ in their mean age of acquisition either.
Research question (P2) combines order of acquisition and frequency. Growth is understood as the age at which the number of types for each class increased most – a definition taken from the notion of vocabulary spurt or vocabulary explosion, a phenomenon describing

that children acquire a large amount of words in a short period of time (e.g, E. V. Clark, 1993; Menyuk, Liebergott, & Schultz, 1995). It is conceivable that this phenomenon may also appear for specific word classes, in this case adjectives. Due to their identical semantic complexity the "explosion" should occur at the same age for Classes 2a and 2b and for Classes 3a and 3b, respectively. As a consequence of the differences in semantic complexity, Class 1 should "explode" earlier than or simultaneously with Class 2. The "explosion" for Class 2 should happen earlier than or simultaneously with Class 3. Class 3 should "explode" earlier than or simultaneously with Class 4.

I predict these outcomes particularly for adjectives that occur together with nouns, i.e., adjectives in attributive or predicative structures, because complexity is defined by means of entailment and gradability properties. Entailment properties determine by which operation adjective and noun are combined. The noun is also relevant for some kinds of gradable adjectives because it can provide the comparison class, which is necessary to calculate the standard of comparison. Therefore, the syntactic structures in which adjectives are produced in spontaneous speech must be considered, too.

If the outcomes of the spontaneous speech study are not as predicted this can have at least five reasons. First, it is possible that the mean age of acquisition or the growth of the adjective classes are not the optimal variables to indicate semantic complexity. Among the two, the mean age of acquisition is expected to yield more reliable results than the growth patterns because for the latter frequency plays a role. Frequency is prone to be influenced by the child's communicative needs, hence whether an adjective or an adjective class is produced more or less often can have reasons that go beyond the child's linguistic abilities. Second, other factors besides semantic complexity may influence the order of acquisition such as properties of the input or how important the adjective is to express the child's communicative needs. The importance for the child is also likely to be reflected in the frequency of use, thus frequency of use by the child may also influence the order of acquisition. Third, the classification according to entailment and gradability may not be appropriate and for instance notional classes are more relevant for investigating children's adjective use. Fourth, the *Semantic Complexity Hierarchy* may not be correct because the proposed definition of semantic complexity must be a different one. Fifth, spontaneous speech data allow only indirect inferences about children's semantic representation. Therefore, it is unclear whether they have the adult-like semantics on which the notion of semantic complexity builds. As a consequence, the comprehension of adjectives should be assessed in a better controlled experimental setting.

In the comprehension experiment in Chapter 5 the following research questions are addressed:

(C1a.) Do monolingual German-speaking children differentiate between absolute and relative gradable dimensional adjectives when defining the standard of comparison?

(C1b.) Does the standard of comparison change with age?

(C2a.) Which gradable dimensional adjectives do monolingual German-speaking children interpret relative to a comparison class?

(C2b.) Do interpretation patterns change with age?

Note that in the comprehension experiment only equally complex adjectives, namely relative (*groß* ('big'), *klein* ('small')) and absolute (*sauber* ('clean'), *dreckig* ('dirty')) gradable dimensional adjectives (Class 2), are investigated. These two classes are interesting for the notion of semantic complexity in acquisition for two reasons. First, they are supposed to be more complex than non-gradable adjectives because they denote relations between individuals and degrees rather than just properties of individuals. Hence, they are a case in point for the question of how children interpret these adjectives and how they figure out that adjectives can be gradable. Second, relative (RG) and absolute (AG) gradable dimensional adjectives are claimed to be equally complex. However, although they have the same semantic type, they are interpreted differently because of lexical semantic properties, e.g., their scale types. With regard to the acquisition of these two classes, one must ask how children come to differentiate adjectives that have the same semantics but differ in their interpretation. Following Snyder (2007), hypothesis H1 predicts that they should be acquired, i.e., interpreted adult-like, at the same age: the standard of comparison should be located at one of the endpoints of the scale for absolute gradable adjectives and around the center of the scale for relative gradable adjectives at the same age (C1a.). The standard for relative gradable adjectives should be defined as a range of degrees resulting in gap between relative gradable antonyms, i.e., a range of sizes which count as neither big nor small. For absolute gradable adjectives, the standard should be a single degree on the scale which is shared by the positive and the negative antonym. Regarding question (C2a.), children should adjust the standard of comparison for relative gradable adjectives according to the comparison class, but should not adjust the standard for absolute gradable ones according to the comparison class. It is open whether the differences in the standard of comparison and in the relevance of the comparison class are acquired together or whether they can be acquired independently.

It is a crucial question whether there is a developmental stage in which children do not have an adult-like interpretation of gradable adjectives (C1b. and C2b.) and what exactly characterizes the non-adult-like interpretation. It is possible that children start with the same interpretation for adjectives that are equally complex, i.e., for all gradable adjectives. "The same interpretation" can show up in two different response patterns.

First, either relative or absolute gradable adjectives may be selected as the default interpretation. For this pattern, it is open whether absolute or relative gradable adjectives are more likely to be the default interpretation. From a learnability perspective both options are conceivable. If AGs are the default interpretation, all adjectives would be interpreted as having closed scales, hence the comparison class would be relevant for any gradable adjective. For example, *big* would be either only the absolute biggest object (in a context) or every object except the smallest object (in a context) and the standard would be determined

independent of the comparison class. In this case the child would have to acquire that not all scales are closed, but that there exist also open scales. What kind of evidence could the child use to acquire the distinction between RGs and AGs? One possible indicator of the distinction between RGs and AGs is the distribution of degree modifiers. *Completely* for instance is only possible with AGs, but not with RGs. To recognize which degree modifiers are licensed in combination with gradable adjectives may help the child to arrive at the target-like distinction (see Syrett, 2007).

If alternatively RGs are the default interpretation, all adjectives would be interpreted as having open scales, hence the comparison class would be relevant for every gradable adjective. For instance, *clean* would not only mean maximlly clean (= without any dirt) and *dirty* would not be true of objects that are only slightly dirty. In this case the child would have to acquire that not all scales are open, but that there exist also closed scales. Again, the child could use the distribution of degree modifiers as evidence to acquire the distinction between RGs and AGs. However, economy assumptions speak against a default interpretation of all gradable adjectives as RGs. If children are conservative learners and follow the law of parsimony, they should select the hypothesis with the fewest assumptions. This would be the option in which the absolute gradable interpretation is the default because the comparison class is not relevant for the interpretation of gradable adjectives. On the other hand, one can also argue that it is the comparison class that distinguishes gradable adjectives and that it is therefore an indicator to acquire the distinction between gradable and non-gradable adjectives as well as the distinction between RGs and AGs.

Second, RGs and AGs may be interpreted alike if neither is interpreted adult-like initially. For this pattern, it is open which semantic representation of gradable adjectives children have instead. It is possible that initially, children have a less complex interpretation of gradable adjectives and interpret them similar to non-gradable adjectives.

These non-adult-like interpretations remind of E. V. Clark's (1973) *Semantic Feature Hypothesis*. E. V. Clark (1973) claims that initially children do not know the complete adult meaning of the words they use. Instead, the lexical entries for words are incomplete "such that these partial entries correspond in some way to some of the features or components of meaning that would be present in the entries for the same words in the adult's lexicon" (E. V. Clark, 1973, p. 72). She further claims that general features should be acquired before specific features. However, Clark notes that it is open what a feature of meaning is. Given the classification I proposed possible features with regard to adjectives are entailment and gradability properties. The order in which these properties are acquired is addressed in Chapter 6.

In this section I presented the research questions of this thesis based on the hypothesis that the order of acquisition of adjectives is determined by their semantic complexity. Previous studies on the acquisition of adjectives proposed alternative determining factors such as properties of the input. In the following chapter I summarize previous research and discuss the findings. I conclude that the input is related to the acquisition of adjectives, but that it does not completely determine the order of acquisition. This is the second

acquisition hypothesis of this thesis.

Chapter 3

Previous research on the acquisition of adjectives

The syntax and semantics of adjectives presented in Chapter 2 hints at a complex acquisition task for the language learner: the child has to discover the syntactic environments in which adjectives can occur and has to acquire the lexical meaning of specific adjectives as well as the compositional semantics of adjectives. This chapter summarizes the main findings of previous acquisition studies concerned with the acquisition task in the production and comprehension of adjectives. The chapter is structured as follows: In Section 3.1, I report on studies investigating the development of the lexicon. This section concerns the emergence of the word class ADJECTIVE compared to other word classes as well as the development of the adjective class in terms of notional classes. In Section 3.2, I report on studies examining attributive and predicative uses of adjectives. In this regard, I also present fast mapping studies investigating the acquisition of novel words embedded in syntactic adjective and noun contexts. Section 3.3 focuses on studies investigating the acquisition of relative gradable dimensional adjectives – Class 2b in the *Semantic Complexity Hierarchy*. In addition, I report on studies comparing the acquisition of relative gradable dimensional adjectives to absolute gradable dimensional adjectives – Class 2a in the *Semantic Complexity Hierarchy* – and to non-gradable adjectives – Class 1 in the *Semantic Complexity Hierarchy*. With regard to hypothesis H1 stating that the acquisition order of adjectives is determined by the adjectives' semantic complexity, I assess whether previous acquisition findings provide evidence that differences between adjective classes are mirrored in acquisition data (Section 3.4). In Section 3.5, I discuss the use of adjectives in the input as an alternative predictor for the acquisition of the semantics of adjectives. I conclude that the input does not completely determine the acquisition order of adjectives (H2).

3.1 Adjectives in the lexicon

Studies investigating children's early lexical development focus on the production of children's first words (e.g., E. V. Clark, 1993; Dromi, 1987; Menyuk et al., 1995; Nelson, 1973). Compared to nouns and verbs, the word class ADJECTIVE received less attention. Like nouns and verbs, adjectives are among the first words children produce. However, adjectives turn out to have a smaller proportion in the mental lexicon and to increase less than nouns and verbs. It is important to note that the classification of single-word utterances as adult part-of-speech categories is difficult because single-word utterances lack morphosyntactic marking. A word of the target language can belong to different parts-of-speech; its reference is determined by the syntactic structure. For instance, the German word *weiß* can refer to an adjective as in *Der Schnee ist weiß* ('The snow is white') or it can refer to the first and third Person singular form of the verb *wissen* ('know') as in *Ich weiß es nicht* ('I don't know'). In children's single-word utterances the same word can be used for different purposes. According to E. V. Clark (1993), an adjective such as *hot* for example can refer to the property of having a high temperature, but can be used as a word for "oven" or "Be careful! Don't touch it!". Various taxonomies using different criteria for the classification of early words have been suggested. In particular, adjectives have been classified in various ways (see Kauschke, 2000). In some studies adjectives and verbs have been classified as PREDICATES (e.g., Gentner, 1982) or as OTHER. In the following, I concentrate on studies that classify adjectives as MODIFIERS (Dromi, 1987; Menyuk et al., 1995; Nelson, 1973). These studies have a similar defintion of the class MODIFIERS. However, different words may have been included in this class.

Nelson (1973, p.17) defines modifiers as "words that refer to properties or qualities of things or events". According to Nelson (1973), this definition holds for attributes, states, locatives and possessives. Nelson (1973) analyzed the first 50 words of 18 English-speaking children until the age of 25 months. The spontaneous speech of six children was recorded from 10 or 11 months, of five children from 12 or 13 months, and of seven children from 14 or 15 months. Words were excluded from the analysis if they were imitations or if form and meaning were not consistent, e.g., using *hot* as as a word for high temperature and as a word for "oven". The words included in Nelson's analysis were classified according to their content or reference as nominals, action words, personal-social words, function words, and modifiers. As critized by Dromi (1987), this coding remains rather intransparent. For the analysis, the first occurrence of each word was counted. Nelson (1973) found that all word classes were among children's first 50 words, but not distributed equally. Across children, the largest class were nominals (65%) followed by action words (13%) and modifiers (9%). Personal-social words (8%) and function words (4%) occurred least often. In contrast to other word classes such as nominals, the proportion of modifiers hardly changed over time.

Similar to Nelson (1973), Dromi (1987, p. 93) defines modifiers as "words that refer to properties and qualities of things and events". However, examples for words that were analyzed as modifiers are lacking. Dromi (1987) examined the speech of one Hebrew-

speaking child during the one-word phase between 10 and 17 months of age. Dromi's (1987) analysis is based on diary notes, but according to the author, it should be evaluated as a regularly and frequently description of the child's early vocabulary and the non-linguistic context rather than an anecdotal recording of words uttered by the child. In addition, 13 audio recordings of 30 to 180 minutes spontaneous speech were analyzed. Words repeatedly used were classified as object words, action words, modifiers, social words, or indeterminant words. Similar to Nelson's (1973) findings for English, all word classes were present in the child's speech, but not distributed equally. Overall, the largest class were object words (59%) followed by indeterminant words (16%) and action words (14%). Social words (7%) and modifiers (4%) occurred least often. In total, the child produced 11 words classified as modifiers. Eight of these words were analyzed as the adult part-of-speech ADJECTIVE. The first adjective was used at 16 months. Indeterminant words are words that could belong to different classes. According to Dromi (1987), the child used in particular adjectives with various reference. For instance, the word *hot* was used for "being hot" or for "heater" or "oven".

Menyuk et al. (1995) uses the same definition of modifiers like Nelson (1973) and Dromi (1987). Menyuk's (1995) analysis of children's early vocabularies is based on diaries taken by mothers until their children's productive vocabularies reached 50 words. The speech of 53 English-speaking children was documented over a period of 30 months, either between birth and 30 months or between 9 and 39 months. Words were excluded from the analysis if they were imitations. The words included in the analysis were classified as nominals, action words, personal-social words, and modifiers. Across children, the largest class were nominals (66%) followed by by action words (18%) and personal-social words (10%). Modifers were the smallest class (6%).

Kauschke (2000) uses a slightly different classification of children's early words. In order to apply the same classes to words in single-word and multiple-word utterances, words were classified as nouns, verbs, adjectives, personal-social words, onomatopoeia, relational words, pronouns, function words, and other words. Kauschke (2000) analyzed transcripts of ten minutes spontaneous speech produced by 32 German-speaking children at 13, 15, 21 and 36 months. Each occurrence of a word was listed in order to count types as well as tokens. The word list contained the target-like form without morphological marking rather than the children's specific realization of the word. Kauschke (2000) found that across children 751 words were produced. 39.3% of the words were nouns, 24% were verbs, and 9.5% were adjectives, 27.2% were personal-social words, relational words, pronouns, function words, or onomatopoeia. The proportion of noun and verb types and tokens increased with age. In contrast, the proportion of adjective types and tokens increased only slightly. The proportion of adjective types increased from 2.38% at 13 months to 6.04% at 36 months. The proportion of adjective tokens increased from 2.18% at 13 months to 3.24% at 36 months.

Tribushinina, Gillis, and De Maeyer (2013) compared the frequency of adjective use of ten 2-year-old and ten 3-year-old Dutch-speaking children. Adjective frequency was

measured as token frequency. In contrast to Kauschke's (2000) results, Tribushinina et al.'s (2013) findings point to a significant increase in the adjective token frequency between age 2 and 3. However, the findings for Dutch may have been different if longitudinal rather than cross-sectional data would have been analyzed.

In summary, findings from previous studies investigating the development of early vocabularies indicate that adjectives emerge in spontaneous speech between age 1 and 2. Studies by Blackwell (2005) and Tribushinina et al. (2014) examined the word class ADJECTIVE in more detail. In these studies, notional adjective classes have been employed to define the concepts that adjectives express. Table 3.1 lists the notional adjective classes used in Blackwell (2005) and Tribushinina et al. (2014).

Table 3.1: Adjective classes in the notional coding of adjectives in Blackwell (2005) and Tribushinina et al. (2014).

Notional class	Examples	Note
AGE	*young, young, old, adult*	
BEHAVIOR	*wild, funny, cautious, talkative, awkward*	Subclass of HUMAN PROPENSITY in Blackwell (2005)
COLOR	*blue, green, colorful*	
CONFORMITY	*other, same, different, normal, correct*	Subclass of OTHER in Blackwell (2005)
DIMENSION	*big, long, narrow, high*	Together with POSITION class coded as SPATIAL PROPERTY in Tribushinina et al. (2014)
MENTAL STATE	*angry, crazy, happy, afraid*	Subclass of HUMAN PROPENSITY in Blackwell (2005)
MODAL	*needed, necessary, forbidden, possible*	Only used by Tribushinina et al. (2014)
ORDINAL NUMBERS	*first, second, sixth*	Only used by Tribushinina et al. (2014)
PHYSICAL PROPERTY	*loud, fast, soft, dirty, dry, round, hot, sweet*	
PHYSICAL STATE	*sick, tired, hungry, dead*	Subclass of HUMAN PROPENSITY in Blackwell (2005)
POSITION	*front, back*	Together with DIMENSION class coded as SPATIAL PROPERTY in Tribushinina et al. (2014)
QUANTITY	*alone, complete, empty*	Only used by Tribushinina et al. (2014)
TEMPORAL PROPERTY	*last, next*	Only used by Tribushinina et al. (2014)
VALUE	*good, fantastic, interesting*	

The PHYSICAL PROPERTY class includes several subclasses. In Blackwell (2005) these are Texture (e.g., *sticky, smooth, rough*), Configuration (e.g., *straight, broken*), Substantiality (e.g., *full, thick*), Sense (e.g., *dark, loud, sour*), Consistency (e.g., *soft, rigid*), Matter (e.g., *wooden, plastic*), Form (e.g., *round, flat, straight*), Speed (e.g., *fast, slow, quick*),

Wetness (e.g., *wet, dry, moist*), Cleanliness (e.g., *clean, dirty*), Appearance (e.g., *pretty, ugly*), and Edibility (e.g., *raw, bitter*).

In Tribushinina et al. (2014) the PHYSICAL PROPERTY class consists of the subclasses Surface (e.g., *rough, smooth, dry, dirty*), Configuration/Functionality (e.g., *open, loose, broken*), Taste/Edibility (e.g., *sweet, salty, raw*), Smell (e.g., *fragrant, smelly*), Sound (e.g., *quiet, noisy*), Shape (e.g., *round, flat, straight*), Light (e.g., *dark, light*), Consistency (e.g., *soft, rigid, thick*), Matter (e.g., *wooden, plastic*), Speed (e.g., *fast, slow, quick*), and Temperature (e.g., *warm, cold*).

In a spontaneous speech study of two English-speaking children between 2;03 to 4;11 and 2;03 to 4;00 years, Blackwell (2005) searched for occurrences of adjectives from a preselected list of 272 adjectives. In total, the children produced 2,963 and 4,299 adjective tokens distributed over 118 and 143 adjective types. Adjectives were regarded as acquired when they occurred with two different nouns or in two different syntactic contexts. Seventy-six adjectives were acquired by both children and were analyzed further. At age 2;04, all main notional classes defined (DIMENSION, COLOR, VALUE, AGE, PHYSICAL PROPERTY, HUMAN PROPENSITY) were found in the children's speech. By age 3, the COLOR class grew as well as subcategories of the PHYSICAL PROPERTY CLASS class. Between age 3 and 4, the two children acquired more specific DIMENSION, PHYSICAL STATE, and BEHAVIOR adjectives. Between age 4 and 5, the children acquired more specific VALUE and MENTAL STATE adjectives. One child used more DIMENSION, PHYSICAL PROPERTY and COLOR adjectives than HUMAN PROPENSITY, AGE and VALUE adjectives. The other child used more DIMENSION, PHYSICAL PROPERTY and HUMAN PROPENSITY adjectives than COLOR, AGE and VALUE adjectives. Tribushinina et al. (2014) extended this line of research to other languages with an open adjective class (Dutch, German, French, Hebrew, Turkish) and to an earlier time window (1;08 to 2;08 years). Unlike in Blackwell (2005), Tribushinina et al. (2014) did not search for adjectives from a preselected list. Rather the children's speech samples were searched for words satisfying the following criteria: (i) being a content word, (ii) being an open-class word, (iii) having adjectival semantics, (iv) having adjectival inflectional morphology, (v) being used in adjectival syntactic positions (Tribushinina et al., 2014). The speech samples varied between children from 1,323 to 9,725 utterances. The children studied produced between 183 and 3014 adjective tokens distributed over 53 to 360 adjective types. The order of emergence of notional classes in the languages they studied was similar to Blackwell's (2005) results for English. PHYSICAL PROPERTY, COLOR, SPATIAL PROPERTY, EVALUATIVE and CONFORMITY adjectives were among the early occuring classes. These classes emerged before INTERNAL STATE, BEHAVIOR, TEMPORAL and MODAL adjectives and ordinal numbers. Regarding the analysis of the token frequency, Tribushinina et al. (2014) used a multilevel model to estimate the growth curves for the different notional classes for each child. Variation in the growth curves were found between notional classes and between children. The probability of use for PHYSICAL STATE adjectives increased steadily with age. With respect to the use of PHYISCAL PROPERTY adjectives no growth was detected. The probability of use for COLOR adjectives hardly

changed for some children, whereas for other children an increase, but at different ages, was observed. The class of SPATIAL and VALUE adjectives showed diverse growth-curves for the ten children. Tribushinina et al. (2014) conclude that some classes (e.g., AGE, BEHAVIOR, MENTAL STATE, TEMPORAL PROPERTIES, MODAL) do not grow much with age. In contrast, other classes (e.g.,COLOR, PHYSICAL STATE, SPATIAL) are used frequently and increase with age.

The findings by Blackwell (2005) and Tribushinina et al. (2014) suggest differences between notional classes with respect to their frequency of use. Moreover, their findings imply that adjectives are acquired in an orderly fashion. The question of which factors determine this order is answered differently by Blackwell (2005) and Tribushinina et al. (2014). Blackwell (2005) concludes that the order of acquisition is affected by the adjectives' notional properties in that adjectives that are independent of the noun meaning – so-called 'categorematic adjectives' – may be used earlier and more frequently than adjectives that are dependent on the noun meaning – so-called 'syncategorematic adjectives'. However, DIMENSION adjectives do not fit this explanation. They are acquired early although their interpretation is dependent on the noun meaning. Tribushinina et al. (2014) conclude that adjectives with so-called 'concrete semantics', i.e., adjectives that refer to immediately perceptible concepts, emerge earlier and are more frequent than adjectives referring to less accessible abstract properties. Moreover, Tribushinina et al. (2014) consider the occurrence of adjectives in child-directed speech as relevant: according to the authors, adjectives that emerge early are highly frequent in the input and important for child-parent interactions.

Tribushinina et al. (2014) found that caregivers used adjectives on average more often than children, but in both child speech and child-directed speech the probability of adjective occurrence increased significantly with the age of the children. Tribushinina et al.'s (2014) findings suggest a correlation between adjective use in child speech and child-directed speech, which decreases with age. However, the adjective use differed between caregivers: for some of the parents adjective use was relatively constant across age, whereas for others adjective use increased significantly with age. Moreover, the correlations between adjective use in child speech and child-directed speech differed across notional adjective classes. Tribushinina et al. (2014) conclude that there is a close relation between children's adjective use and the adjective use of their caregivers such that parents adjust their adjective use to the capacities and interests of the child. However, it is open whether parents or children trigger the increasing adjective use.

Blackwell's (2005) findings for two mother-child dyads point to a small influence of children's adjective use on their mothers' adjective use because adjective use was similar between the two mothers. Moreover, the children produced adjectives from other notional classes than their mothers. On the other hand, Blackwell (2005) found a negative correlation between input properties (input frequency, syntactic diversity, variety in noun co-occurrence) and the age of acquisition. Input frequency and syntactic diversity were significant predictors of the age of acquisition. Nevertheless, Blackwell (2005) concludes that input properties can account only partially for the variance in the age of acquisition

for different adjectives.

In this regard, it is also relevant in which syntactic frames adjectives occur in the input and whether the child can infer the adjectives' meaning from these structures. In Section 3.2, I review studies investigating children's comprehension of adjectives in attributive and predicative structures. In addition, I report on studies examining the production of attributive and predicative structures.

3.2 Attributive and predicative uses of adjectives

Recall that in German adjectives can occur in predicative structures and in attributive structures. In predicative structures (e.g., *Dumbo ist grau* 'Dumbo is gray') the adjective is the complement of a copula and not inflected. In attributive structures (e.g., *der graue Elefant* 'the gray elephant') the adjective is part of the DP and agrees with the noun in number, gender, and case. It is assumed that adjectives in predicative position describe a property of an entity, whereas adjectives in attributive position specify the set of possible referents (e.g., Dixon, 2004). This section reports on studies investigating the development of these two structures in production (Becker, 2000; Bittner, 1998; Blackwell, 2001; Clahsen, Eisenbeiss, & Vainikka, 1994; Eisenbeiss, 2000; Menyuk et al., 1995; Nelson, 1976; Saylor, 2000; Tribushinina, Gillis, & De Maeyer, 2013). In addition, comprehension studies are presented that examine whether children are able to use the syntactic context to infer the meaning of novel words (Klibanoff & Waxman, 2000; Mintz & Gleitman, 2002; Mintz, 2005; Taylor & Gelman, 1988; Waxman & Booth, 2001) and whether they are able to use the adjective to establish reference (Ninio, 2004; Fernald et al., 2010; Thorpe, Baumgartner, & Fernald, 2006).

3.2.1 Attributive and predicative structures in spontaneous speech

To investigate the development of the DP in German-speaking children, Clahsen et al. (1994) analyzed spontaneous speech data of one child between the age of 1;10 and 2;09 years. During this time the child produced 300 noun phrases containing an adjective. The data indicate that adjectives and nouns are combined already between age 1;10 and 2;00. Clahsen et al. (1994) found that at this age, adjectives and determiners occurred in complementary distribution: in the 116 utterances containing an adjective, 113 were of the form adjective + noun, whereas only 3 were of the form determiner + adjective + noun. Utterances of the form determiner + adjective were not produced at this age. The number of determinerless noun phrases containing adjectives decreased with age. Clahsen et al. (1994) conclude that at this age, the child's grammar provides only one structural position for prenominal elements, which can be occupied either by the adjective or by the determiner. The authors also found that the number of agreement errors decreased with age from 80% of agreement errors at age 1;10 until 2;00 to 27% agreement errors at age 2;04 until 2;09. Eisenbeiss (2000) extended this line of research to a larger number of children.

In Eisenbeiss' (2000) spontaneous speech data analysis of seven German-speaking children, four children passed an acquisition stage in which adjectives and determiners occurred in complementary distribution.

The findings by Clahsen et al. (1994) and Eisenbeiss (2000) are supported by Bittner's (1998) findings. Bittner (1998) anaylzed the spontaneous speech of nine girls between 1;11 and 2;10 years. Bittner (1998) also identified a stage in which only one prenominal element was used. If children used an adjective at this stage, the adjective restricted the set of possible referents. As of age 3, children were able to use a determiner as well as an adjective in prenominal position. In addition, Bittner (1998) found that the acquisition of agreement marking, in particular with respect to number, gender, and dative and genitive case, was not complete at age 3.

All three studies focused on attributive adjective structures. Regarding the development of predicative adjective structures, Becker (2000) analyzed the spontaneous speech of four English-speaking children between 2;00 and 2;07 years. Becker's (2000) analysis revealed that in about 50% of the utterances with adjectives in postnominal position the copula *be* was realized, e.g, *this is orange*. In the other 50% of the utterances the copula was missing, e.g., *this empty*.

In conclusion, the findings for the development of attributive structures in German and predicative structures in English suggest that both structures develop between age 2 and 3 (see also Tribushinina, Gillis, & De Maeyer, 2013, for Dutch). In what follows, I present studies comparing the acquisition of attributive and predicative structures. Menyuk et al. (1995) analyzed 100 consecutive multimorphematic utterances of 53 English-speaking children at 25, 29, and 35 months. Menyuk et al. (1995) found structures that contained a modifier and a noun. This structure is similar to the attributive use of adjectives. These structures constituted 10% of the multimorphematic utterances at 25 months, 8% at 29 months, and 5% at 35 months. However, it is important to note that utterances of that type were not restricted to adjectival modifiers as in *little ball*, but also included demonstratives as in *this ball*. In addition, Menyuk et al. (1995) found structures that contained a noun and a copula followed by an adjective or a noun, e.g., *Dumbo is {gray/an elephant}*. Only if the copula is followed by an adjective, this structure corresponds to the predicative use of adjectives. Structures of the form noun + copula + adjective/noun constituted 10% of the multimorphematic utterances at 25 months, and 14% at 29 and 35 months. These data indicate that at age 2;01, both constructions that could contain an adjective were equally distributed. However, at 2;11 copular constructions were more frequent than modifiers in prenominal position. However, this conclusion is limited because the structures were not restricted to adjectives.

Nelson (1976) studied the syntactic position of adjectives and the correponding discourse functions in the spontaneous speech of English-speaking children at 2;00 and 2;06 years based on the first 100 utterances of 24 transcripts. Nelson found that younger children with a lower MLU (1.0 - 2.5 morphemes) produced predicative adjectives more frequently than attributive adjectives (66.7 vs. 33.3%), whereas older children with a higher MLU (2.5

- 4.5 morphemes) produced predicative adjectives less frequently than attributive structures (23.3 vs. 75.3%). Adjectives expressing so-called 'object-states' (e.g., *cold, dirty, open*) and 'animate states' (e.g., *hurt, dead*) were used most frequently in predicative position. They mostly served to comment on referents. Adjectives referring to descriptive properties (e.g., *big, white*) were used frequently in attributive position with the function to further specify referents. Overall, attributive adjectives were more frequent than predicative structures (59.6 vs. 40.1%).

Blackwell (2001) found that attributive and predicative adjectives emerged simultaneously in the spontaneous speech of the three English-speaking children (age 1;6 to 5;1 years) she studied. Across the 6,507 multi-word adjective utterances analyzed, attributive uses were more frequently attested than predicative uses (62 vs. 30%). This distribution held across age. In contrast to Nelson (1976), Blackwell (2001) found that attributive adjectives were used as well to comment on already established referents. Blackwell (2001) argues that the adjectives' properties in terms of 'time-stability' explain the data better than the discourse functions of attributive and predicative adjectives. According to this approach, nouns are time-stable because the entities nouns denote remain the same over time. In contrast, verbs are time-instable because they express rapid change. Adjectives can encode either time-stable properties or time-instable properties. In the data analyzed by Blackwell (2001), time-stable properties such as COLOR and DIMENSION were mainly produced in attributive position; on average in 48% and 74% of the utterances. In contrast, less time-stable properties such as HUMAN PROPENSITY adjectives (e.g., *sick, happy*) were used primarily in predicative structures, on average in 63% of the utterances. PHYSICAL PROPERTY adjectives, a large and heterogeneous class (see Section 3.1), occurred equally frequent in both positions, on average in 33% and 38% of the utterances, depending on how time-stable the respective property is.

Saylor (2000) analyzed the use of predicative and attributive adjectives of three English-speaking children between 2;03 and 5;00 years. These children produced between 954 and 2,054 adjective tokens. For two children, the adjective tokens were distributed equally across predicative and attributive structures. One child used slightly more attributive than predicative adjectives (56 vs. 39%). This distribution was consistent across age. In addition, Saylor (2000) found a relation between syntactic position, time-stability of the property denoted by the adjective, and notional class similar to the one observed by Blackwell (2001).

In a study by Tribushinina, Gillis, and De Maeyer (2013) (see Section 3.1), between age 2 and 3 the syntactic complexity of structures containing adjectives increased from adjectives used in single-word utterances to adjectives used in target-like structures. Already at age 2, both attributive and predicative structures were attested, but attributive adjectives were more frequently used than predicative adjectives.

To summarize, previous acquisition findings suggest that during the third year of life noun phrases become more complex. In particular, two prenominal elements, e.g., a determiner and an adjective, can be used instead of either of them. Moreover, from age 2,

copular verbs are realized in utterances containing adjectives pointing to a development of these structures between age 2 and 3. With regard to the distribution of predicative and attributive structures, previous results are inconclusive. The conflicting results may be partly due to different codings. Menyuk's (1995) structures that resemble predicative and attributive adjective structures are also possible with other word classes in the position of the adjective. It is possible that the frequencies of these other word classes skew the picture with respect to the distribution of adjective structures. Nelson (1976) coded single-word utterances consisting of an adjective as predicative, whereas Blackwell (2001) and Tribushinina, Gillis, and De Maeyer (2013) excluded these utterances when comparing attributive and predicative structures. The inclusion of single-word adjective utterances may explain the higher proportion of predicative adjectives compared to attributive adjectives found by Nelson (1976) in younger children. Independent of the inconclusive results regarding the distribution of adjective structures, the findings by Nelson (1976) and Blackwell (2001) point to a relation between the syntactic position of the adjective and the property denoted by the adjective.

3.2.2 Attributive and predicative structures in comprehension studies

Previous studies used different morphosyntactic structures to investigate whether children are able to use them as a cue for inferring the meaning of novel words (Klibanoff & Waxman, 2000; Mintz & Gleitman, 2002; Mintz, 2005; Taylor & Gelman, 1988; Waxman & Booth, 2001). A series of studies examined whether children are able to recognize that a novel word in a syntactic noun context. e.g., *This is a tiv*, refers to an object category and that a novel word in a syntactic adjective context, e.g., *this is a tiv one* or *this one is tiv*, refers to an object property.

In a fast mapping experiment, Taylor and Gelman (1988) tested 32 English-speaking children with a mean age of 2;02 years. The experiment consisted of a naming phase and a test phase. In the naming phase children were either presented with four familiar toys, e.g., two stuffed dogs and two stuffed birds, or with four novel toys, i.e., two kinds of unfamiliar toys. One toy of each category was made out of pale green fake four, the other toy was made out of bright yellow and black plaid material. In the syntactic noun context the experimenter named one of the four toys six times with for instance *This is a tiv.* In the syntactic adjective context the experimenter named one of the four toys six times with for instance *This is a tiv one.* In the test phase children were asked to perform five actions on the four objects, e.g., put in a box, hide, hold in one hand, put on your head, hug, put on a table, throw in the air, point to, or give away. In the syntactic noun context the experimenter asked the child for instance *Can you put a tiv in the box?.* In the syntactic adjective context the experimenter asked the child for instance *Can you put a tiv one in the box?.* In addition, the experimenter asked for each of the four objects whether it is *a tiv* and a *a tiv one*, respectively. Children who selected in at least four of five trials the named object or the toy from the same category were scored as having a category interpretation of

the novel word. Children who selected in at least four of five trials the named toy or the toy made from the same material were scored as having a property interpretation of the novel word. The category interpretation was predicted for children who heard the novel word in a syntactic noun context, the property interpretation was predicted for children who heard the novel word in a syntactic adjective context. The results indicate that children interpret novel words used as nouns as a label for the object category if it referred to an unfamiliar object, i.e., they selected an object from the same category as the named object more often than an object made from the same material. Children who heard novel words used as adjectives did not have a preferred interpretation. Contrary to prediction, they did not select an object of the same material more often than an object of the same category.

Waxman and Booth (2001) extended this line of research to younger children. The responses of 34 children with a mean age of 1;02 years were included in the analyisis. Similar to Taylor and Gelman's (1988) study, children heard novel words in a syntactic noun context or in a syntactic adjective context. However, in Waxman and Booth's study the adjectives had an adjective-like morphological marking, e.g., *blik-**ish***, and they were presented in predicative rather than in attributive structures. For each novel word children received two familiarization trials, a contrast trial, and a property or a category test trial. In the first familiarization trial children were presented with two objects from the same or from different basic-level categories having the same color, e.g., two purple horses or a purple bear and a purple lion. In the syntactic noun context the experimenter uttered the following description: *These are blickets. This one is a blicket and this one is a blicket.* In the syntactic adjective context the description was: *These are blikish. This one is blikish and this one is blikish.* These descriptions were repeated in the second familiarization trial. In the contrast trial the experimenter presented one of the purple objects and an object from a different category with a different color, e.g., a purple horse and an orange carrot. The orange carrot was described as *This is not {a blicket/blikish}*. The purple horse was described as *This one is {a blicket/blikish}*. In the test phase the child saw either a purple horse and a blue horse (property trial) or a purple horse and a purple plate (category trial). The experimenter presented again a purple horse saying *This is a blicket* or *This one is blikish* and then asked with respect to the two test objects *Can you give me the blicket?* or *Can you give me the blikish one?*. In the category trials children were expected to select the familiar object, i.e., the purple horse, more often than the unfamiliar object, i.e., the purple plate, when asked *Can you give me the blicket?*. In the property trials they should perform at chance level when asked *Can you give me the blicket?*. When asked *Can you give me the blikish one?* children's choices should be reversed. Waxman and Booth (2001) found that children who were asked *Can you give me the blicket?* selected the familiar object more often in category trials (68%) than in property trial (44%). When asked *Can you give me the blikish one?* children selected the familar object equally often in category trials (50%) and in property trials (52%). Waxman and Booth (2001) conclude that at age 1;02, children hearing a novel word in syntactic noun contexts interpret it as a label for the object category. Children hearing a novel word in syntactic adjective contexts have

no preferred interpretation of the word as a label for an object property even if the novel word has an adjective-like morphological marking. Waxman and Booth (2001) did not find a difference regarding children's interpretations of novel words depending on whether the objects in the familiarization trials were from the same or from different basic-level categories than the test objects.

In contrast, Klibanoff and Waxman (2000) observed differences in children's interpretation of novel words within and across basic-level categories. Their study examined children's ability to map novel words in syntactic adjective contexts to object properties. For this purpose, Klibanoff and Waxman (2000) investigated under which conditions this interpretation is prompted. Similar to the novel words used in Waxman and Booth's (2000) study, the novel words had an adjective-like morphological marking. The object descriptions in Klibanoff and Waxman's study differed from the ones used by Taylor and Gelman (1988) and Waxman and Booth (2001) because the novel word was presented in an attributive adjective structure with a basic-level count noun rather than with the indefinite pronoun *one*. The novel word was presented in a syntactic adjective context only. Klibanoff and Waxman (2000) tested 64 English-speaking 3- and 4-year-old children. In each trial, the experimenter presented a target object, e.g., a bumpy green horse or a bumpy green rhinoceros. Then, the experimenter presented two test objects either from the same or from a different basic-level category. Both objects had a different color than the target object. One object matched the target property, i.e., it was also bumpy, the other object contrasted the target object, i.e., it was smooth. In the same basic-level condition children saw for instance a bumpy purple horse and a smooth purple horse after they saw the bumpy green horse. In the different basic-level condition children saw a bumpy yellow horse and a smooth yellow horse after they saw the bumpy green rhinoceros. The experimenter introduced the target object as *This is a very blikish horse* and as *This is a very blikish rhinoceros*, respectively. When presenting the two horses the experimenter asked the child for *a horse that is blikish*. Klibanoff and Waxman (2000) found that overall 4-year-olds selected the object with the matching property more often than the 3-year-olds (90 vs. 62%). However, also the 3-year-olds selected the object with the matching property significantly above chance level, but only if the target object was from the same basic-level category than the test objects. The 4-year-olds selected the matching test object also if the target object was from a different basic-level category. This finding suggest that in the process of acquiring that novel adjectives refer to object properties children initially extend the property to objects within a basic-level category and later to objects across basic-level categories.

Mintz and Gleitman (2002) also investigated under which conditions children are able to interpret novel words as object properties. In their study the novel words modified a basic-level count noun in attributive as well as in predicative adjective structures. In contrast to Klibanoff and Waxman's (2000) study, three target objects which shared a property rather than a single target object were presented, e.g., a striped cup, a striped horse, and a striped block. The novel word for that property was presented without an adjective-like morpholgical marking, e.g., *stoof*. The experimenter described each target

object as *Look at this stoof {cup/horse/block}. This {cup/horse/block} is very stoof.* After the presentation of all three target objects the experimenter repeated *These things are all stoof.* The two test objects were for example a boat with stripes and a cup with stars. This way, one test object shared the property of the target objects and one test object shared the category of one of the target objects. The children were asked to select *the stoof one* and to tell *which of the things is stoof.* Mintz and Gleitman (2002) found that both 2- and 3-year-old English-speaking children selected the object with the same property as the target objects more often than the object from the same category as one of the target objects. Findings by Mintz (2005) support this result. In addition, using the same task as Mintz and Gleitman (2002), Mintz (2005) showed that 2- and 3-year-old children interpret novel words as object properties when the novel word modified the superordinate-level noun *toy* or when it modified novel nouns referring to unfamiliar objects. In contrast, when the novel word modifed an underspecifed noun such as *one* or *thing*, only 3-year-olds, but not 2-year-olds, had a preference for the property interpretation.

In summary, previous findings indicate that children as young as age 1;02 interpret novel words in syntactic noun contexts as a label for the object category. As of age 2, children interpret novel words presented in syntactic adjective contexts (attributive and predicative structures) as object properties rather than object categories. At age 2, children are also able to extend the novel word to properties of objects from a different category. Moreover, previous findings indicate that structures in which the novel word modifies a specific basic-level count noun are beneficial to prompt this interpretation. In addition, multiple exemplars are necessary to illustrate the property denoted by the novel word. These observations raise the question whether adjectives are presented this way also in natural settings and not only in experimental settings. In other words, are adjectives presented in these beneficial contexts in the input?

Excursus: Adjective structures used in the input

Findings by Blackwell and Olson (2008), Davies, Lingwood, and Arunachalam (2019), and Sandhofer and Smith (2007) suggest that adjectives occur in a variety of structures, which may not all be beneficial to recognize that adjectives refer to object properties rather than object categories. Blackwell and Olson (2008) analyzed 6,106 maternal utterances containing adjectives that modified nouns and pronouns provided to three children between 1;06 and 4;11 years. In 56% of the maternal utterances the adjective modified referentially specific nouns (e.g., *This is a cute dog*), which have been claimed by Mintz and Gleitman (2002) to help children mapping a novel word to an object property. This means in half of the utterances children received input that it is not helpful to infer the meaning of adjectives because the modified element was an underspecified noun or a pronoun (e.g., *This is a blue one/thing*). Sandhofer and Smith (2007) analyzed the input provided to 12 children between 2;00 and 2;06 years in situations that were conducive to talk about properties of things. In these situation parents played with their children with teddies, cups, and stacking rings that differed in color, size, and texture. This way, it was likely

that the parents make comparisons within basic-level categories (e.g., fluffy vs. scratchy bear) and between basic-level categories (e.g., the red bear with its red cup). The parents used adjectives 4,785 times in one of the following structures:

(3.1) a. This is a red bear.
b. This is a red one.
c. This is red.
d. This is the red.

Example (3.1a) indicates most unambiguously that *red* is an adjective because it modifies a basic-level count noun. This structure was used in 24% of the utterances. In example (3.1b) the noun is replaced by an indefinite pronoun. Mintz (2005) showed that this structure makes it more difficult for children to identify *red* as an adjective. This structure was used in 20% of the utterances. The structure in (3.1c) is ambiguous because the position after the copula could also be filled by a proper name. This structure was used in 30% of all utterances. Example (3.1d) is not an optimal cue to understand that *red* is an adjective because it is unambiguously used in a noun context. This structure was used in 6% of all utterances. In addition, adjectives were used in 20% of all utterances as single-word utterances.

Davies et al. (2019) argue that adjectives in predicative position and adjectives with a contrastive use should be more useful for the acquisition of adjectives than adjectives in attributive position and adjectives with a descriptive use. When the adjective is in attributive position, the property denoted by the adjective must be remembered until the noun has been heard. In contrast, when the noun is uttered before the adjective as in predicative structures, the child's focus is on the intended referent and the adjective can be interpreted relative to the noun. Contrastively used adjectives as in *That bag is heavy and this bag is light*, but not descriptively used adjectives (*This bag is heavy*), signal the dimension where the two referents of the same object category differ. This may help the child to map the meaning of the adjective to the relevant dimension (Davies et al., 2019). Davies et al. (2019) analyzed input data from three different sources (during free play, from the text provided in children's books, and from the spontaneous speech produced during shared book reading) and found a mismatch between the adjective contexts that are theoretically helpful for the acquisition of adjectives and the adjective contexts that children hear in the input.

In conclusion, in the input provided to children adjectives occur in structures different from the ones used in experimental settings. Moreover, it is likely that the input includes structures that are not optimal to infer the meaning of the adjective.

The studies described so far in this section focused on the question whether children are able to interpret novel words as object properties when the words are presented in predicative and attributive adjective structures. In addition, a short note on adjective structures

in the input addressed the question whether parents use these structures when talking about object properties. In order to master this task the language learner must map a single word to a relevant property. Fernald et al. (2010, p. 192) claim that this is a process "different from the process involved in rapidly interpreting familiar adjective-noun phrases in continuous speech." This process involves the integration of the adjective and the noun meaning in attributive structures in order to restrict the set of possible referents.

Ninio (2004) investigated the comprehension of attributive adjective structures in Hebrew-speaking children between 1;06 and 4;04 years. In each trial of the experiment the experimenter presented four pictures displaying familiar objects. The objects were from two different categories and had two different properties, e.g., a big and a small teddy, and a big and a small clock. Because adjectives appear postnominal in Hebrew, the test prompts used were of the form *Give me the N-Adjective*. Ninio (2004) found that in 41% of the trials children did not select the correct picture. The most frequent error was to select an object from the correct catgeory, but with an incorrect property, e.g., children selected the big teddy when asked for the small teddy. Ninio (2004) concludes that children have problems with integrating the adjective and the noun meaning, and therefore they do not process the adjective. Three questions remain open. First, it is unclear whether the ability to integrate the adjective and the noun meaning improves with age because the age range of the children investigated is very large. Second, Ninio (2004) used a variety of adjectives with different notional and semantic properties.[1] Semantic properties have consequences for the combination of adjective and noun meaning. It is possible that children's difficulty with the task is due to difficulty with the interpretation of specific adjectives. Third, it is open whether children aquiring languages with a different word order exhibit the same difficulty with integrating adjective and noun meaning, which results in the same error patterns.

The first and the third question were addressed by Thorpe et al. (2006) using a design similar to Ninio's (2004). Thorpe et al. (2006) tested English-speaking children divided into three age groups (1;06-1;10, 2;04-2;08, 3;05-4;06 years). The youngest age group did not select the correct picture in 54% of the trials, the middle age group in 40% of the trials, and the oldest age group in 15% of the trials. Regarding children's errors, Thorpe et al. (2006) found in the 2-year-old group errors similar to the ones reported by Ninio (2004), i.e., the children selected an object from the correct category mentioned in the test prompt, but with an incorrect property. In addition to this error, another error was observed in the youngest and oldest age group. These children selected an object with the correct property mentioned in the test prompt, but from an incorrect category, e.g., a yellow fish when asked for a yellow balloon. The findings by Thorpe et al. (2006) indicate that children's ability to integrate adjective and noun meaning improves with age. Moreover, the same error patterns as in the Hebrew experiment were found in the English experiment. To find out whether children's responses result from not processing the adjective, Fernald et al.

[1] The following adjectives were used: *white, black, green, yellow, red, blue, small, big, short, long, full, empty, clean, dirty, awake, asleep, happy, sad, naked, dressed, whole, cut.*

(2010) used an online looking-while-listening procedure with children aged 2;06 and 3;00. Fernald et al. (2010) examined whether and at which point in an attributive structure children use the adjective to find the correct referent. Their study was restricted to color adjectives in the test prompts. Test prompts such as *Can you find the blue car?* were used in three different conditions. In the 'same color/different objects' condition children saw for instance a blue car and a blue house. In this condition, the adjective was not informative for the identification of the correct referent. In the 'different colors/different objects' condition children saw for instance a blue car and a red house. In this condition, the adjective is informative, but not necessary to identify the correct referent. In the 'different colors/same object' condition children saw for instance a red car and a blue car. In this condition, the adjective is the only cue to identify the correct referent. If children are able to use the adjective as a cue, they should look faster to the target picture in the 'different colors/different objects' and in the 'different colors/same object' condition. Moreover, if they are able to integrate the adjective and the noun meaning, children should continue looking to the target picture when they hear the noun. Fernald et al.'s (2010) findings suggest that at 2;06, children do not take advantage of the adjective in the 'different colors/different objects' and in the 'different colors/same object' condition to establish reference more quickly – they rather wait to hear the noun. Nevertheless, 2;06-year-old children are able to use the adjective to find the correct referent. Hence, they are able to integrate the adjective and the noun meaning. However, Fernald et al. (2010) observed more individual differences between children at age 2;06 than at age 3. At age 3, children were able to interpret the adjective incrementally, i.e., they looked to the target before they heard the noun.

In summary, previous findings on the interpretation of attributive adjective structures suggest that integrating the adjective and the noun meaning to identify the correct referent is a gradual process that is mastered between 2;06 and 3;00 years. It is possible that word order plays a role in mastering this task: children acquiring a language in which the adjective follows the noun may master it later than children acquiring a language in which the adjective precedes the noun. Because previous studies tested adjectives from a single adjective class or it was not controlled for the semantic properties of the adjectives used, it is open whether the findings can be generalized across all types of adjectives.

3.3 Gradable adjectives

Studies on the acquisition of compositional and lexical semantic properties are scarce. One exception is research on the acquisition of relative gradable dimensional adjectives – Class 2b in the *Semantic Complexity Hierarchy*. Section 3.3.1 reviews studies on the production and comprehension of these adjectives. In Section 3.3.2, studies comparing the acquisition of relative gradable adjectives with the acquisition of other adjective classes are reported.

3.3.1 Relative gradable dimensional adjectives

As illustrated in Chapter 2, gradable adjectives denote relations between individuals and degrees on a scale that refers to a specific dimension, e.g., height. Gradable adjectives must be interpreted relative to a standard of comparison. For relative gradable adjectives, the standard of comparison is calculated depending on the context – the comparison class. As a consequence, the standard of comparison for relative gradable adjectives can shift from context to context. Thus, the same object judged for instance *big* in one context could be judged *little* in another context. Previous acquisition studies focused on three different aspects regarding the interpretation of relative gradable dimensional adjectives. Early studies investigated the acquisition of adjectives referring to different dimensions and to the different poles of scales (e.g., Bartlett, 1976; Eilers et al., 1974; Klatzky, Clark, & Macken, 1973). A large body of research examined whether children have a so-called 'relational' or a so-called 'nominal' interpretation of relative gradable adjectives (e.g., Panzeri & Foppolo, 2012; Sera & Smith, 1987; Smith et al., 1986; Tribushinina, 2013). In addition, previous studies investigated to which differences in context children are sensitive and whether they adjust the standard of comparison accordingly (e.g., Barner & Snedeker, 2008; Ebeling & Gelman, 1988, 1994; Gao, Zelazo, Sharpe, & Mashari, 2014).

Eilers et al. (1974) presented 14 English-speaking children between 2;06 and 3;06 years pairs of objects. The objects of each pair differed either in size, length, or width. This way, Eilers et al. (1974) examined the comprehension of *big-little*, *long-short*, and *wide-narrow*. The experimenter presented a pair of objects and asked the child for example *Give me the big one*. Each object pair was presented twice: once with a positive pole adjective in the test prompt and once with a negative pole adjective. In this forced-choice task, children made fewer errors for *big-little* than for *long-short*, and fewer errors for *long-short* than for *wide-narrow*. In addition, children made more errors for the respective positive pole adjective than for the negative pole adjective. In a follow-up experiment, the number of errors for positive and negative pole adjectives did not differ. Eilers et al. (1974) conclude that more specific dimensions such as length or width are acquired later than more general dimensions such as size. Moreover, their findings do not suggest a difference in the acqusition of positive pole and negative pole adjectives.

Bartlett (1976) extended Eilers et al.'s (1974) research to younger and older children between 2;01 and 4;10 years. The authors added object pairs that differed in height to the object pairs differing in size, length, or width. Similar to Eilers et al. (1974), Bartlett (1976) found that adjectives referring to more general dimensions are acquired before adjectives that refer to more specific dimensions. At age 5, children's performance of adjectives such as *wide-narrow* was not yet at ceiling. With regard to the acquisition of positive and negative pole adjectives, Bartlett (1976) found that polarity affected children's performance only for specific dimensions: at age 2, no difference between positive and negative pole adjectives that refer to overall size, i.e., *big* and *little*, was observed. For adjectives referring to more specific dimensions, less errors were found for positive pole adjectives than for negative

adjectives. However for *wide-narrow*, this pattern was reversed for some age groups.

In a word learning experiment, Klatzky et al. (1973) taught English-speaking children aged 3;09, 4;02, and 4;08 years novel words corresponding to the positive and negative pole adjectives *big-small*, *high-low*, *long-short*, and *thick-thin*. Across age, it took children more trials to learn the novel words corresponding to the negative pole adjectives (on average 3 to 6 trials depending on the dimension) than to learn the novel words for the positive pole adjectives (on average 2 trials). In addition, children made more errors for the novel words corresponding to the negative pole adjectives (on average 20 to 40% depending on the dimension) than for the novel words corresponding to the positive pole adjectives (on average 10 to 20% depending on the dimension).

Although it was not the focus of the studies by Bartlett (1976), Eilers et al. (1974), and Klatzky et al. (1973), their findings imply that at age 2, children are able to compare two objects with regard to some dimensions and to decide on the basis of this comparison for which object an adjective is true. According to Sera and Smith (1987), this ability can be regarded as a first instance of a relational interpretation of relative gradable dimensional adjectives. However, it has been claimed that children start with a so-called 'nominal' interpretation of relative gradable adjectives, i.e., that children interpret relative gradable adjectives as "having extent" (H. H. Clark, 1970). Possible evidence for this assumption is provided by Panzeri and Foppolo (2012) and Tribushinina (2013).

Panzeri and Foppolo (2012) tested 20 3-year-old and 20 5-year-old Italian-speaking children. Panzeri and Foppolo (2012) presented objects in isolation (e.g., a wooden rod). According to the authors, this presentation does not provide any standard of comparison. The objects used did not have a special intended use, hence participants were prevented to apply a functional standard. Because no other objects were present in the visual context, participants could not use a perceptual standard. In a truth-value judgement task the participants had to judge whether the test sentence, e.g., *This is tall*, is true of the object presented. Such test prompts provide no other contextual standard. Because no standard was available, the participants were expected to answer *I don't know* or show a chance distribution of accepting and rejecting the test sentence. Contrary to prediction, the 3-year-olds accepted the test sentence in 88% of the trials and the 5-year-olds in 60% of the trials. These results point to a non-relational interpretation of relative gradable adjectives. According to Panzeri and Foppolo (2012), children did not have a 'yes'-bias because they rejected test sentences in the filler trials. However, it is crucial to note that this experimental setting cannot exclude that children may have some conceptual representation of the objects used. It is possible that this conceptual representation provided the standard of comparison (see Ebeling and Gelman (1988) on children's normative standards and Sedivy, Tanenhaus, Chambers, and Carlson (1999) on adults' standards in incremental processing of relative gradable adjectives). What is more, findings by Huang and Snedeker (2013) from an online visual-word study indicate that 5-year-olds looked to the target object more rapidly when hearing test prompts such as *Point to the big coin* if the display contained two coins rather than one coin. This finding demonstrates that at age 5, children

typically evaluate whether a property denoted by a relative gradable adjective is true for an object by comparing it to another object.

So-called 'extreme' interpretations of relative gradable adjectives have been claimed to be another reflex of children's nominal interpretations (Smith et al., 1986; Tribushinina, 2013). This means children apply the adjective only to the object which displays the respective property most. For instance, *big* is applied only to the biggest object, and *small* only to the smallest object in a series of objects. If the relative gradable adjective is interpreted as "having extent" or "having the property denoted by the adjective", the 'extreme' objects can be viewed as the best exemplars representing the respective property. Tribushinina (2013) observed this interpretation in the responses of 2- and 3-year-old Dutch-speaking children. The children saw series of seven objects increasing in size. The objects were either prototypically big (elephants, hippos, houses, planes), prototypically small (mice, chickens, gnomes, babies), or prototype-neutral (balloons, cakes, monkeys, umbrellas). Each series was accompanied by the question *Which X do you find {big/small}?*, e.g., *Which houses do you find big?*. In more than 50% of the responses the children selected only the biggest object when asked for *big*, and only the smallest object when asked for *small*. Note that the way of asking – a subjective question – may have favored answers in which the participants chose only the best exemplar. A similar pattern has been described by Smith et al. (1986) for English-speaking 3-year-olds' interpretation of *high* and *low*. In Smith et al.'s study, the task was different because the height of a single object was manipulated and the child was asked for each position *Is this {high/low}?*.

Another case in point for children's nominal or relational interpretation of relative gradable adjectives is the ability to shift the standard of comparison according to the context, i.e., the comparison class. Several studies investigated the kind of contextual information that influences children's calculation of the standard of comparison (Booij & Sassoon, 2014; Barner & Snedeker, 2008; Ebeling & Gelman, 1988, 1994; Gao et al., 2014; Sera & Smith, 1987; Smith et al., 1986; Tribushinina, 2013). Sera and Smith (1987) investigated whether 2- to 4-year-old English-speaking children are able to relabel objects as *big* and *little* depending on the visual context. This means the standard of comparison had to be shifted according to changes in the perceptual context. The perceptual context was created from three circles of different sizes: one small circle, one medium circle, and one big circle. The differences in size between the small, medium, and big circle varied across trials. In the first part of each trial, the experimenter presented for example the small and the medium circle and the child had to label the two circles *big* and *little*. In the second part of each trial, the small circle was removed and replaced by the big circle. The child again had to label the two circles *big* and *little*. Hence, the medium circle should have been labeled *big* in the first part and *little* in the second part. The results indicate, that 2- and 3-year-olds indeed have difficulty to shift their judgements for *big* and *little* depending on the perceptual context. However, children's performance increased when the size differences between the objects were large. Given large size differences between objects, children shifted the labels above chance-level. This result indicates a developing relational interpretation of relative

gradable adjectives.

Ebeling and Gelman (1988) investigated whether children use normative standards to determine whether an object is big or little in addition to the standard calculated based on the perceptual comparison class. A normative standard is calculated based on a conceptual comparison class, i.e., based on the conceptual representation of the object kind and its typical size. Ebeling and Gelman (1988) used three different tasks to test 2-, 3-, and 4-year-old English-speaking children's production and comprehension of *big* and *little*. The test stimuli were created from nine mittens differing in size. In the normative task the children saw the second largest and the second smallest mitten one at a time and should judge whether the specifc mitten is big or little. All three age groups performed above chance in this task indicating that at age 2 children have a normative standard for familiar objects. The production task was similar to the task used in Sera and Smith (1987). Again, all three age groups performed above chance pointing to a standard calculated based on the perceptual comparison class. However, in trials in which the perceptual standard conflicted with the normative standard, children had greater difficulty to calculate a perceptual standard. In these trials children had to label the bigger of two small mittens *big* or the smaller of two big mittens *little*. Hence, the normative standard seemed to influence the perceptual standard in the production task. The comprehension task was analogous to the production task except that the child was asked to give the experimenter *the big one* or *the little one*. In this task, all age groups selected the correct object more often than predicted by chance and performed better than in the production task.

Ebeling and Gelman (1994) investigated a third kind of standard: a so-called 'functional' standard, which is calculated based on the intended use of an object. The normative and the perceptual production task were similar to the ones in Ebeling and Gelman (1988) except that various kinds of clothes rather than only mittens were presented. In the functional task the experimenter presented for example a hat and a doll and asked the child *Here's a doll and here's a hat for the doll. Is this hat big for the doll or little for the doll?*. The relevant trials were those in which children should first calculate a normative and then either a perceptual or functional standard that conflicted with the normative standard. For instance, the child saw a small hat first and should label it *little*. Then, the child saw a small hat and a tiny hat, or a small hat on a tiny doll and should relabel the hat *big*. The results indicate that at age 3, children are able to switch from a normative to a perceptual standard (76% correct responses) and from a normative to a functional standard (69% correct responses) more often than predicted by chance. The finding was replicated in a follow-up experiment with 2-year-olds. In contrast to these findings, 2- and 3-year-olds did not shift their standard when they had to calculate a perceptual standard first and then had to switch to a normative standard.

In the studies by Ebeling and Gelman the visual context differed depending on whether children had to calculate a normative, perceptual, or functional standard. Smith et al. (1986) and Tribushinina (2013) examined the effect of normative and perceptual comparison classes for the same visual context. As described above, Tribushinina (2013) used

prototypically big, prototypically small and neutral objects. Smith et al. (1986) used prototypically low (bunnies) and prototypically high (birds) objects. Both studies observed that 3-year-olds calculated the standard of comparison independent of their knowledge about the object category. From age 5, knowledge about the object category influenced children's calculation of the standard. For instance, the standard for *big* or *high* was higher for prototypically big or high objects than for prototypically small or low objects. These findings indicate that children are able to shift their standard according to the comparison class in the sense that they can use multiple reference points, namely preceptual and normative ones.

In Gao et al.'s (2014) sorting task, perceptual and normative comparison classes played a role as well. In addition, the comparison class was linguistically encoded by the modifed noun. Hence, in contrast to most previous studies, Gao et al. (2014) used a task in which also the explicit comparison class was manipulated. Gao et al. (2014) tested 24 3- to 4;05-year-old and 24 4;06- to 6;06-year-old English-speaking children. Two stimulus sets were used in the experiment. One set consisted of two water noodles, two hockey sticks, two crayons, and two marbles. The other set consisted of two baseball bats, two golf clubs, two chalks, and two jacks. According to Gao et al. (2014), two kinds of objects in each set were normatively big (water noodles and hockey sticks, baseball bats and golf clubs) and two kinds of objects were normatively little (crayons and marbles, chalks and jacks). The two objects from the same basic-level category also differed in size, e.g., there was a big and a small hockey stick. In the first part of the experiment, children saw the big and the little version of the same basic-level category, e.g., the big and the small hockey stick, and were asked to put the big hockey stick in one place and the little hockey stick in another place. In the second part, the big hockey stick was replaced by the big crayon, i.e., the child saw a normatively big toy and a normatively small toy. The experimenter asked the child to put the big toy in one place and the little toy in another place. As a consequence, depending on whether the child was asked to sort the hockey sticks or the toys, the hockey stick that was present in both parts should have been judged as little in the first part and as big in the second part. Except for two children, all children switched their judgements between the parts in at least six out of eight trials. In the third part of the experiment, the child saw all eight objects from one stimulus set together, e.g., a big and a small water noodle, a big and a small hockey stick, a big and a small crayon, and a big and a small marble. First, the child was asked to sort the objects according to the superordinate-level category "toy". To illustrate this way of sorting the experimenter put the hockey sticks in the *big*-place and the crayons in the *little*-place. The child was expected to complete the sorting by putting the water noodles in the *big*-place and the marbles in the *little*-place. Second, the child was asked to resort the objects such that the big members of each basic-level category should have built one group and the little members of each basic-level category should have built the other group. To illustrate this way of sorting the experimenter put the big hockey stick and the big crayon in the *big*-place and the little hockey stick and the little crayon in the *little*-place. The child was expected to complete the sorting by putting the big water noodle

and the big marble in the *big*-place and the little water noodle and the little marble in the *little*-place. In the group of 3- to 4;05-year-old children, 18 of 24 children were not able to resort the objects in the way explained by the experimenter. Eleven of these children were younger than 4 years. Gao et al. (2014) conclude that children up to age 4 are not able to shift their standard of comparison for the same perceptual context, although the comparison class was encoded linguistically by a basic-level or superordinate level noun, and despite the experimenter's explanation of the expected sorting.

In order to exclude children's normative standards for specific kinds of objects, Barner and Snedeker (2008) used unfamiliar objects with novel names in four experiments. In each experiment they tested 16 to 32 4-year-old English-speaking children. In Experiment 1 children saw nine novel objects called 'pimwits' ranging in height from 1 to 9 inches. The nine objects were presented all at once in pseudo-randomized order. In the *tall*-trial the children were asked to find *the tall pimwits*, in the *short*-trial to find *the short pimwits*. The average height of the smallest object selected in the *tall*-trial was 7.19, the average height of the tallest object selected in the *short*-trial was 3.19. This difference points to different standards of comparisons for *tall* and *short*. In an additional task, children had to judge which of a 1 inch and a 3 inch pimwit is taller and which of a 7 inch and a 9 inch pimwit is shorter. These comparative judgements were available to 15 out of 16 children. In Experiment 2 the perceptual context changed because four pimwits were added to the nine original objects. The small 'distractors' reduced the mean size of objects, the tall 'distractors' increased the mean size of objects. Compared to Experiment 1, children had different standards of comparison for *tall* and *short*. In the short distractor condition, the standard of comparison was lower in Experiment 2 than in Experiment 1 for *tall* ($m = 5.44$) and for *short* ($m = 2.19$). In the tall distractor condition, the standard of comparison was higher in Experiment 2 than in Experiment 1, but only for *tall* ($m = 8.44$). These findings indicate that 4-year-old children are sensitive to changes in the perceptual context and the distributional properties of the object array when calculating the standard of comparison. Experiment 3 was identical to the short distractor condition of Experiment 2 except that the distractors were of a different kind. They looked different and had a different name ('tulvers'). The standards for *tall* ($m = 6.89$) and *short* ($m = 2.94$) differed from the standards in Experiment 2, but not from the standards in Experiment 1. This result indicates that children attend to kind information when they determine the comparison class and calculate the standard of comparison. In Experiment 4, the perceptual context was the same as in Experiment 3, but the distractors were also called 'pimwits' instead of 'tulvers'. The standard for *tall* ($m = 5.69$) differed from the standard in Experiment 3, but not from the standard in Experiment 2. This finding indicates that children attend to linguistic information porvided by the noun to determine the comparison class and to calculate the standard accordingly. However, for *short*, the standard was not adjusted to changes in the explicit comparison class. On the basis of these between-subjects comparisons, Barner and Snedeker (2008) conclude that at age 4, children are able to use the information provided by the noun to determine the comparison class and to calculate the standard according to

the distributional properties of that comparison class.

Barner and Snedeker (2008) manipulated the mean height of the objects in the visual array and found that differences in the mean height result in different standards of comparisons. Booij and Sassoon (2014) also manipulated the distributional properties of the object array, but in a way that did not affect the mean size of objects. Booij and Sassoon (2014) tested 8-year-old Dutch-speaking children. The children saw four different sets of 12 pencils and where asked to put the big pencils in a basket. For each set, the smallest pencil was 5cm long and the longest pencil was 16cm long. The mean length of pencils was always 10.5cm, but the standard deviation differed between stes. In the first distribution the pencil length increased linearly, in the second distribution there were two clusters of pencils with a gap in the middle, the third distribution was similar to the first distribution but had two leaps[2], and the fourth distribution had two small clusters of pencils at the extremes and one large cluster in the middle. Children's choices of big pencils show that the standard of comparison was a fixed length (around 11cm) in all four distributions. This finding indicates that 8-year-old children have a standard of comparison for *big* that is not influenced by distributional properties different from the mean length.

The studies reported in this section focused on the comprehension of relative gradable dimensional adjectives. Hohaus et al. (2014) investigated the emergence of gradable adjectives in spontaneous speech and the use of gradable adjectives in comparison constructions. Hohaus et al. (2014) do not specify whether they focus only on relative gradable dimensional adjectives and whether they expect differences between different classes of gradable adjectives. However, to my knowledge Hohaus et al. (2014) are the first to propose developmental stages for the acquisition of gradable adjectives that take compositional semantic properties into account. The authors observed that different structures containing gradable adjectives were acquired in a specific order in the spontaneous speech of the three German-speaking children investigated. First, adjectives in the unmarked positive form were acquired in the first half of the third year of life followed by contextual comparatives (e.g., *Hans ist 1.70m groß.* ***Maria ist größer.*** 'John is 1.70m tall. Mary is taller.'[3]) at the end of the third year of life. Between 3 and 4 years, the pronominal measure phrase construction (e.g., *Hans ist 1.70m groß.* ***Bill ist genauso groß.*** 'John is 1.70m tall. Bill is that tall, too.') and superlatives were acquired. As of age 6, phrasal *than*-constituents (e.g., *Maria ist gößer als Hans.* 'Mary is taller than John.'), overt measure phrase constructions (e.g., *Hans ist genau 1.70m groß.* 'John is exactly 1.70m tall.'), and degree questions (e.g., *Wie alt ist Maria?* 'How old is Mary?') were acquired. Given that the comparison constructions found in German are not universal (Beck et al., 2009), Hohaus et al. (2014) conclude that this order of acquisition for gradable adjectives and comparison constructions is based on parametric variation. The logic behind that conclusion is the following:

[2] The pencils in the first distribution had the following lengths (in cm): 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16. In contrast, the pencils in the third distribution had the following lengths (in cm): 5, 5.5, 8, 8, 9, 10, 11, 12, 13, 13, 15.5, 16.

[3] Because Hohaus et al. (2014) provide German examples from children only for some constructions, I use fictional examples to illustrate the constructions.

assume that parameter A and parameter B determine whether two specific constructions containing a gradable adjective are possible in a language and that there are languages that have the positive setting of parameter A and the negative setting of parameter B. Then, constructions indicative of parameter B should not be acquired before constructions indicative of parameter A. Regarding the semantic knowledge associated with the different constructions and the acquisition stages at which they occur, Hohaus et al. (2014) propose that initially, children have a semantics of gradable adjectives that does not include degrees. According to Hohaus et al. (2014), a degree semantics can be assumed as soon as children use gradable adjectives in pronominal measure phrase constructions, contextual comparatives, and superlatives, i.e., approximately between age 3 and 4.

In summary, the findings of previous studies on relative gradable dimensional adjectives – Class 2b in the *Semantic Complexity Hierarchy* – indicate that children acquire relative gradable adjectives that refer to more general dimensions such as size before those that refer to more specific dimensions such as width. With regard to the question whether positive and negative pole adjectives are acquired differently, the results are inconclusive. The relational interpretation of relative gradable adjectives develops from age 2. As of age 4, children are able to calculate the standard of comparison according to implicit comparison classes, e.g., the perceptual context, and according to explicit comparison classes, i.e., the noun modifed by the adjective. The use of comparative constructions provides first evidence that at age 4 children have a target-like semantic representation of relative gradable dimensional adjectives. In Section 3.3.2, I review previous research on the acquisition of other adjective classes that are part of the *Semantic Complexity Hierarchy*.

3.3.2 Relative gradable dimensional adjectives in comparison to other adjective classes

Few studies compared the acquisition of different adjective classes. In a series of studies, Syrett and her colleagues compared the comprehension of relative gradable and absolute gradable dimensional adjectives – Class 2a in the *Semantic Complexity Hierarchy* (see also Foppolo & Panzeri, 2013). Tribushinina and Gillis (2012) investigated the use of relative and absolute gradable dimensional adjectives. The production and comprehension of color adjectives that belong to the class of non-gradable adjectives – Class 1 in the *Semantic Complexity Hierarchy* – have been explored by Deutsch and Pechmann (1982) and Wagner, Storage, Dobkins, and Barner (2013). Emotion adjectives, which may belong to the class of evaluative adjectives – Class 4 in the *Semantic Complexity Hierarchy* – have been investigated with regard to the question whether they are interpreted as relative or absolute gradable (Britton et al., 2017).

Recall that for relative and absolute gradable adjectives – Class 2a and 2b in the *Semantic Complexity Hierarchy* – the standard of comparison is calculated differently. For relative gradable adjectives, which have open scales, the standard of comparison is context-dependent and is located around the center of the corresponding scale. In contrast,

for absolute gradable adjectives, which have closed scales, the standard of comparison is not context-dependent and located at one of the endpoints of the corresponding scale. Using two different tasks (Presupposition Assessment Task, Scalar Judgement Task) Syrett (2007), Syrett, Bradley, Kennedy, and Lidz (2006) and Syrett et al. (2010) investigated whether 3- to 5-year-old children have different standards of comparison for relative and absolute gradable and whether they distinguish relative and absolute gradable adjectives with respect to their context-dependency. In the Scalar Judgement Task 3- to 5-year-old children saw a series of seven objects (cubes, rods, containers, disks) that displayed the same property but to different extents. Participants had to judge whether the property denoted by the adjective in question is true of each of the objects in the series. For each object a child was asked *Is this ADJ?*. *ADJ* was either a relative gradable adjective (*big, long*) or an absolute gradable adjective (*full, straight, spotted, bumpy*). As expected, children's standard for relative gradable adjectives was located around the center of the scale, i.e., children answered 'yes' with respect to the three biggest or longest objects more often than for the three smallest or shortest objects. Children's standard for absolute gradable adjectives was located at one of the enpoints of the scale, but only for the minimum absolute gradable adjective *spotted*: children gave a 'yes'-response to all objects with a non-zero degree of the property, i.e., having at least one spot. Regarding the maximum absolute gradable adjective *full*, contrary to prediction, children accepted containers as full that were not maximally full, suggesting a non-endpoint standard. These findings were replicated in the Presupposition Assessment Task. In this task, children saw two objects and the experimenter made the request *Give me the ADJ one*. In two of three conditions children should reject the request if it contained an absolute gradable adjective because the two objects violated the uniqueness or the existence presupposition of the definite determiner. This is, either both of the objects could have been described by the adjective, e.g., two maximally full containers, or none of the objects could have been described by the adjective, e.g., two filled but not maximally full containers. In case the request contained a relative gradable adjective, children should always accept it because there is always one object that is big or long relative to the other object of the pair. From age 3, children performed as expected for relative gradable adjectives and minimum absolute gradable adjectives. However, for maximum absolute gradable adjectives children accepted the request even if the existence presupposition was violated. This finding points to a context-dependent interpretation of maximum absolute gradable adjectives. Syrett et al. (2010) argue against a context-dependent interpretation of maximum absolute gradable adjectives. They claim that the apparent context-sensitivity of absolute gradable adjectives is not part of the adjective's lexical meaning as it is in relative gradable adjectives. Instead, absolute gradable adjectives can be used in a context-sensitive way, which results in imprecise meanings. Foppolo and Panzeri (2013) extended the Scalar Judgement Task to antonym pairs such as *full-empty, clean-dirty*, or *big-small*. Their results for Italian-speaking 3- and 5-year-old children are comparable to Syrett et al.'s findings for English. By using negative and positive pole adjectives, it can be analyzed whether children divide the scale completely

between the antonym pairs or whether they exhibit a 'gap' between the negative and the positive antonym (see Section 2.2.3). For *clean-dirty* the results indicate more strongly than for *full-empty* that children do not exhbit a gap between absolute gradable antonyms. However, because participants were presented with either the positive or the negative antonym, it is not possible to compare their individual standards for positive and negative adjectives.

As described in Chapter 2, relative and absolute gradable adjectives also differ regarding the degree adverbs they license. Due to the different scale structures of relative and absolute gradable adjectives only absolute gradable adjectives can be modified by degree adverbs such as *almost* or *completely*. Non-gradable adjectives cannot be modified by degree adverbs at all. Tribushinina and Gillis (2012) investigated whether 2- to 7-year-old Dutch-speaking children have knowledge about the different gradability properties of adjectives based on children's use of degree modifiers in spontaneous speech in combination with non-gradable, absolute gradable, relative gradable and evaluative adjectives. The latter were not further distinguished regarding the scale type (open vs. closed). With respect to the emergence of the adjective classes investigated, Tribushinina and Gillis (2012) found that all adjective classes were attested in spontaneous speech from age 2 onwards. For each of the ten two-year-olds (age 1;11 to 2;03 years) 150 utterances from one recording were analyzed. Across children the class of non-gradable adjectives consisted of one type and 6 tokens, the class of absolute gradable adjectives consisted of 9 types and 23 tokens, the class of relative gradable adjectives included 3 types and 5 tokens, and the class of evaluative gradable adjectives consisted of 4 types and 7 tokens. Tribushinina and Gillis (2012) found that already at age 2 children never modified non-gradable adjectives with degree adverbs, but gradable adjectives. This results points to a distinction of gradable and non-gradable adjectives from age 2. As of age 4, the proportion of adjectives taking degree adverbs in each semantic category was similar to the proportion found in the speech of adults. Based on these quantitative results and the correct use of degree markers, Tribushinina and Gillis (2012) conclude that children distinguish between different classes of gradable adjectives from age 4.

Syrett (2007) investigated whether children use the distribution of degree adverbs to infer the meaning of novel adjectives. Thirty-three English-speaking children between 2;04 and 2;08 years were assigned two one of three conditions. Syrett (2007) used a preferential looking paradigm with a task similar to the one used by Waxman and Booth (2001) described in Section 3.2.2. In the familiarization phase children saw two objects that showed an absolute gradable property, e.g., being transparent, and a relative gradable property, e.g., being tall. The two objects were described by a novel adjective, e.g., *wuggin*. Depending on the condition, the adjective was modified by *very* or *completely* or no degree adverb was used. If children use the adverb to infer the meaning of the novel word, they should infer that *wuggin* means 'tall' when modified by *very*. In contrast, if *wuggin* is modified by *completely*, children should infer that it means 'transparent'. If no adverb was present, children should not have a preference for one meaning unless they have a default interpre-

tation for novel adjectives. In the contrast phase one of the tall and transparent objects was contrasted with a short opaque object. In the test phase children saw a tall opaque object and a short transparent object and were asked *Which one is wuggin?*. The proportion of looks towards the object with the relative property, e.g., the tall object, in the test phase was analyzed. The proportion of looks to the object with the relative property before and after the test question was compared. The results indicate that in the condition in which the novel word was modified by *completely* children looked to the object with the relative property before the test question was uttered and looked away from the object after the test question. In contrast, in the condition in which the novel word was modified by *very* children looked away from the object with the relative property before the test question was uttered and looked to the object after the test question. The findings point to different interpretations of novel adjectives depending on the degree adverb that modified them. When no adverb was present, children did not show a difference in proportions of looks to the object with the relative gradable property before and after the test question was uttered. This finding does not suggest a default interpretation of novel adjectives.

The interpretation of relative gradable dimensional adjectives compared to the interpretation of non-gradable adjectives was investigated by Wagner et al. (2013). To investigate children's interpretation of *tall* and *blue*, Wagner et al. (2013) used the same task as Barner and Snedeker (2008) described in Section 3.3.1. Contrary to prediction, the children between the ages of 2 and 7, with the exception of the 4-year-olds, interpreted color terms relative to a comparative class. Similar to relative gradable adjectives they changed their judgements for *blue* depending on the objects in the perceptual context. According to Wagner et al. (2013), this result suggests that children initially have the default reading that adjectives are dependent on a comparison class. Note, however, that the experimental design involved the visualization of color adjectives as gradable, which may have affected children's interpretation.

Deutsch and Pechmann (1982) examined which adjectives 3-, 6-, and 9-year-old Dutch-speaking children use in referential descriptions. In the task used by Deutsch and Pechmann (1982) children saw an array of eight objects and were asked to describe the object they want the experimenter to select. The objects differed in object category, in size, and in color. As a consequence, the children had to modify the noun by a non-gradable color adjective and a relative gradable size adjective to describe the object the experimenter should select unambiguously. In the group of 3-year-olds 87% of the first descriptions were ambiguous, in the group of 5-year-olds 50% of the first descriptions were ambiguous, and in the group of 9-year-olds 22% of the first descriptions were ambiguous. This means some information was missing. Deutsch and Pechmann (1982) found that the noun was mentioned in most of the descriptions. Children mainly omitted information regarding the object's color, size, or both. Especially in younger children, the relative gradable size adjective was more often omitted than the non-gradable color adjective.

Britton et al. (2017) investigated whether children interpret emotion adjectives, which may belong to the class of gradable evaluative adjectives, as relative or absolute. Britton et al. (2017) presented 4- and 5-year-old English-speaking children graded picture sequences of people with happy, sad, and angry faces. The children's task was to find all the happy/sad/angry faces. The children saw either nine faces varying in the intensity of the expression, or nine faces varying in the intensity of the expression plus four additional faces with maximal intensity of the expression. Britton et al. (2017) found that children shifted their standards for *happy, sad*, and *angry* depending on the number of faces in the perceptual context. These findings are similar to children's standard adjustments for *tall* and *short* found by Barner and Snedeker (2008) (see Section 3.3.1). Britton et al. (2017) conclude that children interpret emotion adjectives as relative gradable.

Nelson and Benedict (1974) also compared different types of adjectives but they used a different classification of adjectives than the one expressed by the *Semantic Complexity Hierarchy*. The authors classified adjectives as relative (e.g., *tall, long*), absolute (e.g., *leafy, furry*), and contrastive (e.g., *happy-sad, clean-dirty*). In a picture-choice task the adjectives were presented in the positive unmarked form (e.g., *Point to the tall one*) and in the comparative form (e.g., *Point to the taller one*). The choices of the 33 4-, 5-, and 6-year-old English-speaking children were analyzed with respect to response accuracy and latency. Nelson and Benedict (1974) found significantly more errors for the comparative forms than for the positive forms of adjectives. The most errors occurred for relative adjectives. Children took longer to point to a picture if the test prompt contained the comparative form of the adjective than if it contained the positive form. However, this difference was significant only for absolute and contrastive adjectives. Based on children's response accuracy, Nelson and Benedict (1974) conclude that absolute and contrastive adjectives are acquired before relative adjectives. In addition, the findings suggest that unlike other adjectives relative adjectives are processed in a comparative sense.

3.4 Are semantic differences between adjective classes mirrored in acquisition data?

The studies mentioned in Section 3.1, 3.2, and 3.3 have revealed important insights into the acquisition of adjectives. In this section, I summarize the relevant previous findings for the production and comprehension of adjectives regarding the question whether the semantic differences between adjectives described in Chapter 2 are mirrored in the acquisition of adjectives. Moreover, I identify the aspects in the acquisition of adjectives that are still open.

3.4.1 Production of adjectives

Previous studies showed that adjectives are among children's first words (e.g., Kauschke, 2000). Regarding the development of the syntactic structure of utterances containing adjectives, previous studies found that adjectives are first used in single word utterances. Around age 2, adjectives are used in attributive and predicative structures. Attributive and predicative adjective structures emerge around the same age (Blackwell, 2001; Tribushinina, Gillis, & De Maeyer, 2013). Because previous studies coded attributive and predicative structures differently, the findings regarding the frequency distribution of these two syntactic structures are inconclusive.

With respect to the development of meaning properties of adjectives, previous studies focused on the emergence of notional properties such as COLOR, PHYSICAL PROPERTY, HUMAN PROPENSITY etc. (Blackwell, 2005; Tribushinina et al., 2014). The findings suggest that notional properties of adjectives emerge in a specific order across children and across children acquiring different languages. It is open why adjectives emerge in that order because Blackwell (2005) and Tribushinina et al. (2014) provide only tentative explanations. Moreover, these studies investigated the global use of adjectives, i.e., they did not distinguish the syntactic structures in which adjectives occur. This can be relevant because Blackwell (2001) and Nelson (1976) observed a relation between notional properties and syntactic structure, namely that some notional classes tend to occur more frequently in predicative than in attributive structures, whereas other notional classes tend to occur more frequently in attributive than in predicative structures. It is open whether this preference is related to the order of acquisition. It may be that adjectives that favor for instance the attributive position occur earlier in that position than in predicative position.

Previous research on the acquisition of compositional and lexical semantic properties of adjectives has focused mainly on selected aspects such as the use of degree adverbs (Tribushinina & Gillis, 2012) or the use of different comparison constructions (Hohaus et al., 2014). Children's use of degree adverbs and of comparative constructions indicates that they have a semantic representation of gradable adjectives as denoting relations between individuals and degrees on a scale and that they distinguish between different adjective classes at age 4. However, Tribushinina and Gillis (2012) and Hohaus et al. (2014) conclude that the acquisition of the semantics of (gradable) adjectives does not need to be complete by a age 6. A comprehensive picture regarding the development of the compositional and lexical semantic properties of adjectives described in Chapter 2 is missing.

Previous studies investigated either the acquisition of notional properties or the acquisition of semantic properties. Hence, it is open how the acquisition of these properties is related. This is a crucial question which may reveal which compositional and lexical semantic properties underlie the concepts expressed by notional classes.

3.4.2 Comprehension of adjectives

Research investigating the interpretation of novel words indicates that as of age 2, children interpret novel words presented in attributive and predicative adjective structures containing a specific basic-level noun as properties of objects (Mintz & Gleitman, 2002). Two-year-old children are also able to extend the novel word to properties of objects from a different category (Mintz, 2005).

Studies presenting real-world adjectives in attributive structures showed that between 2;06 and 3;00 years children acquire the integration of the adjective and the noun meaning to identify the correct referent (Fernald et al., 2010). It is open whether this ability is mastered at different ages depending on whether the adjective precedes or follows the noun in the target language. Furthermore, it needs to be investigated whether the semantic properties of the adjective affects the ability to integrate the adjective and the noun meaning.

Previous studies on relative gradable dimensional adjectives found that words that refer to more general dimensions such as size are acquired before those that refer to more specific dimensions such as width (Bartlett, 1976; Eilers et al., 1974). With regard to the question whether positive and negative pole adjectives are acquired differently, the results are inconclusive.

A large body of research investigated how children claculate the standard of comparison for relative gradable dimensional adjectives. Regarding the standard degree, Tribushinina (2013) showed for relative gradable dimensional adjectives that 3-year-old children attach *big* and *small* only to the extremes of the range of objects in the visual context, i.e., to the smallest or biggest object. Panzeri and Foppolo (2012) found similar so-called 'nominal' interpretations for 3-year-old Italian-speaking children. Starting around age 4, the standard is context-dependent, or 'relational', and located around the midpoint of the scale. However, Ebeling and Gelman (1988) showed that already at age 2, children can shift the standard for relative gradable dimensional classes by taking conceptual and perceptual comparison classes into account. According to Barner and Snedeker (2008), at age 4, children use the information provided by the noun to determine the comparison class rather than information from the visual context and that they shift the standard according to the noun. In contrast, the children in Gao et al. (2014) interpreted adjectives relative to the noun only if the visual context also changed.

Few studies compared the interpretation of relative and absolute gradable dimensional adjectives (Foppolo & Panzeri, 2013; Syrett, 2007; Syrett et al., 2006, 2010). These findings suggest that at age 3, children distinguish between relative and absolute gradable dimensional adjectives in that the standard for absolute gradable adjectives is located at the endpoint, whereas the standard for relative gradable adjectives is located around the center of the scale. However, the absolute gradable adjective *full* was not always interpreted as having a maximum standard. Because these studies used only positive pole adjectives, it could not be examined whether children divide the scale completely between the antonym

pairs or whether they exhibit a 'gap' between the negative and the positive antonym. First insights come from a study by Foppolo and Panzeri (2013) on Italian. They extended the set of adjectives tested to antonym pairs such as *full-empty* and *clean-dirty*. Their results are comparable with Syrett et al.'s (2006) findings for *full-empty*; for *clean-dirty* the results indicate more strongly that children do not exhbit a gap between absolute gradable antonyms. However, because participants were presented with either the positive or the negative antonym, it is not possible to compare their individual standards for positive and negative adjectives. Syrett (2007), Syrett et al. (2006), and Syrett et al. (2010) found a difference between the interpretation of relative and absolute gradable adjectives also with respect to the relevance of the comparison class: RGs were interpreted relative to a comparison class, whereas AGs were not. In their studies the adjective did not modify a noun, hence the comparison class had to be inferred from the context, i.e., it was implicit.

Although previous studies provide first evidence that children distinguish between absolute and relative gradable dimensional adjectives, several aspects regarding the standard of comparison and the comparison class are still open. Table 3.2 gives an overview over the results and the open aspects of the studies summarized in this section. Taken together the following questions are open: How do children interpret relative and absolute gradable antonymous adjectives? How do children interpret relative and absolute gradable adjectives when the comparison class is linguistically encoded by the noun, i.e., when it is explicit? In other words, can children use the information provided by the noun to determine the comparison class? And for which kinds of gradable adjectives does the comparison class influence the standard of comparison?

Table 3.2: Summary of relevant previous studies on the comprehension of gradable adjectives.

Study	L1	Age	Results	Open aspects
Tribushinina (2013)	Dutch	2-7 years	For *big, small*: • 2- and 3-year-olds relate the adjectives only to the extremes of the scale • From age 5, children start to combine world knowledge and visual context	• Comparison with AGs
Barner & Snedeker (2008)	English	4 years	For *tall, short*: • Children shift the standard according to the visual array • Children use the information provided by the noun to determine the comparison class	• Comparison with AGs • Younger children • Different distributions of objects in the relevant contexts • Within-subjects comparison • Existing objects
Gao et al. (2014)	English	3-6 years	For *big, little*: • 3- and 4-year olds interpret adjectives relative to the noun if the visual context is different, but not if the visual context was the same	• Comparison with AGs
Syrett (2007), Syrett et al. (2006, 2010)	English	3-5 years	For *big, long, spotted, full, straight, bumpy*: • Children have different standards for AGs and RGs • Children shift the standard for RGs, but not for AGs	• Explicit comparison classes • Negative pole adjectives

Note. L1 = First language. All participants were monolingual speakers of the respective language.

3.4.3 Summary

In summary, previous acquisition studies provide first evidence that differences between adjectives in notional as well as in compositional and lexical semantic properties are mirrored in the acquisition of adjectives.[4] I argued that differences in compositional and lexical semantic properties result in differences in semantic complexity. The role of semantic complexity in the acquisition of adjectives has not been examined yet. Among other factors, e.g., the input, semantic complexity is one possible predictor for the order of acquisition of adjectives. To address this question, adjectives used by children in spontaneous speech

[4] These differences between adjective classes have also been detected in online and offline tasks with adults (Aparicio, Kennedy, & Xiang, 2018; Belke, 2001; Frazier, Clifton, & Stolterfoth, 2008; Hansen & Chemla, 2017; Heller & Chambers, 2014; Liao & Meskin, 2015; Pechmann, 1989; Rips & Turnbull, 1980; Ziegler & Pylkkänen, 2016).

should be analyzed according to their compositional and lexical semantic properties. With regard to the interpretation of adjectives, relative and absolute gradable adjectives are a case in point. In light of the compositional properties of adjectives, the influence of explicit comparison classes on the interpretation of these adjectives should be investigated. In addition, more research on positive and negative adjectives is needed to explore their relation with respect to the standard of comparison. Moreover, data from German-speaking children can contribute to a better understanding of the acquisition of adjectives across different languages. The aim of this thesis is to provide an acquisition path for the semantics of adjectives that is based on previous findings and extended by the findings presented in Chapter 4 and 5.

3.5 The role of the input in the acquisition of the semantics of adjectives

Recall the general research question (Q) of this thesis: How do monolingual German-speaking children acquire the semantics of adjectives? Previous research on the acquisition of adjectives provides first evidence that the acquisition of the semantics of adjectives proceeds in a temporal order. In particular, longitudinal spontaneous speech studies found that notional semantic classes are acquired in orderly fashion (see Blackwell, 2005; Tribushinina et al., 2014). However, the explanations for why this particular order is observed remained rather vague and tentative. In contrast, the classification of adjectives and the definition of semantic complexity I proposed in Chapter 2 is based on theoretical considerations regarding the semantics of adjectives. Therefore, it is possible to test the hypothesis formulated in Chapter 2 (H1) that children acquire the semantics of adjectives in an order determined by semantic complexity. Are there other factors that could determine the order of acquisition?

There is evidence that for the acquisition of the lexicon the input is relevant; more than it is for the acquisition of rule-governed linguistic domains such as syntax. Thus, it has been proposed that children acquire the semantics of adjectives driven by the input (Blackwell, 2005; Tribushinina et al., 2014), i.e., by generating abstract knowledge regarding the meaning of adjectives via generalizing across specific occurrences. Nevertheless, the evidence with respect to the role of the input for the acquisition of adjectives is mixed. Tribushinina et al. (2014) found a positive correlation between the frequency of adjective use in children and their caregivers. The authors conclude that caregivers adjust their adjective use to the capacities and interests of their children. However, the correlation between child and parental adjective use decreased with age. This decrease is unexpected given the fact that adjective acquisition was not complete because some notional adjective classes were still missing in the spontaneous speech of the children investigated. If the input is necessary for the acquisition of adjectives, it should be provided throughout the acquisition process. Moreover, the correlation between child and parental adjective use did not show the same pattern across notional adjective classes. If the input is necessary for

the acquisition of notional semantic properties of adjectives, it should be available across notional adjective classes. Besides the frequency of adjective use, Blackwell (2005) takes syntactic diversity and the variety in noun co-occurrence in the input into account. She found a negative correlation between each of these three input properties and the age of acquisition. The factors frequency and syntactic diversity were also able to predict the age of acquisition. However, a general issue regarding the relation between child and parental speech is the question whether children speak in a way their parents did before or whether the way parents speak is a reaction to their children's behavior. Findings by Blackwell (2005) indicate that maternal adjective use is similar across mothers, and hence less likely to be just a reaction to their children's adjective use. Nevertheless, child adjective use can at least be partially influenced by maternal use because the children used adjectives from different notional classes than their mothers (see also Tribushinina et al., 2014). Therefore, Blackwell (2005) concludes that the adjective use in the input is not the only factor determining the acquisition of adjectives. Findings by Blackwell and Olson (2008), Davies et al. (2019), Hohaus et al. (2014), and Sandhofer and Smith (2007) support this conclusion. The results of Hohaus et al.'s spontaneous speech study showed that the order in which children acquire different comparison constructions is not identical to the order of these constructions with respect to their frequency in the input. Blackwell and Olson's as well as Sandhofer and Smith's examinations of parental input revealed that adjectives are produced infrequently. In addition, the syntactic structures in which adjectives are produced by caregivers are often ambiguous as the examples from Sandhofer and Smith (2007) in (3.1) illustrated.

In summary, I assume that the input plays a role in the acquisition of the mental lexicon, and hence for the acquisition of adjectives: the child must hear a form in order to attach a meaning to it. Moreover, the acquisition of the lexicon shows more individual variation than for instance the acquisition of sentence structure, which may be explained by the influence of the input. However, I do not assume that the input is the major factor that determines the acquisition order of adjectives. Therefore, the second hypothesis for the acquisition of the semantics of adjectives is the following:

(H2) The order of acquisition of adjectives is not completely determined by the input.

Unlike previous studies, which investigated mostly notional semantic classes, I investigate the acquisition of compositional and lexical semantic properties implemented in the *Semantic Complexity Hierarchy*. Hypothesis H2 is tested with regard to the effect of the input frequency on children's production of adjectives in spontaneous speech. The following research questions are addressed in the production study reported in Chapter 4:

(P3) Is adjective frequency in child and parental speech correlated?

(P4) Is adjective frequency in parental speech reflected in the adjectives' age of acquisition?

(P5) Are adjectives that children acquire late missing in parental speech before the age of acquisition?

Research question (P3) addresses the general influence of adjective use in the input on children's adjective use. According to hypothesis H2, it is possible that adjective frequency in child and parental speech are positively correlated suggesting a relation between child and parental adjective use. On the other hand, it is also possible that type and token frequency are distributed differently across adjective classes and child and parental speech indicating an independent adjective use in child and parental speech. Research question (P4) addresses the influence of adjective use in the input on the acquisition order of the semantic classes proposed in Chapter 2. According to hypothesis H2, it is possible that the more frequent an adjective occurs in parental speech the earlier it is acquired by the child, but this negative correlation is not necessarily predicted by H2. Moreover, it may also be that frequency and semantic complexity coincide (see Brown & Hanlon, 1970, on syntactic complexity and input frequency). In this case, frequency does not need to determine the acquisition order. Hypothesis H2 further predicts that parents do not necessarily adjust their adjective use to their child's knowledge. Hence, it is possible that parents produce adjectives although these adjectives are not yet acquired by the child (P5).

If the input determines the acquisition of the semantics of adjectives not completely, one must assume that the language learner enters the acquisition process with a predisposition to acquire the semantics of adjectives. This predisposition may come in form of the *Semantic Complexity Hierarchy*, which is triggered by positive evidence from the input. I discuss this topic in Chapter 6.

Chapter 4

Study 1: Adjectives in spontaneous speech

The study reported in this chapter gives an overview over the adjective use in child German.[1] This study provides a more comprehensive picture of the development of adjectives than previous studies (Blackwell, 2005; Hohaus et al., 2014; Tribushinina et al., 2014; Tribushinina & Gillis, 2012) because it takes notional properties as well as various compositional and lexical semantic properties into account. I show that notional classes can be transferred into the semantic classes proposed in Chapter 2, and thus address the crucial question of which compositional and lexical semantic properties underlie the concepts expressed by notional classes. In contrast to previous studies, which investigated only the global use of adjectives (Blackwell, 2005; Tribushinina et al., 2014), I considered the syntactic position in which the adjectives occurred because this may affect the order of acquisition. The focus of this chapter is the question whether the acquisition order of adjectives is determined by their semantic complexity. Moreover, the role of the input for the acquisition order of adjectives is addressed.

For these purposes, spontaneous speech data were analyzed. Spontaneous speech data have several advantages over parental questionnaires or elicited production experiments. First, spontaneous speech data show the real use of words or structures because they are collected in natural settings. Hence, the data are not affected by task-effects (Dromi, 1999; Kauschke, 2000; Stromswold, 1998). Second, spontaneous speech data, if collected with sufficient frequency and duration, yield a detailed picture of the investigated phenomenon and are well suited to detect developmental trends in longitudinal studies (Demuth, 1998; Dromi, 1999; Stromswold, 1998). In addition, findings for spontaneous speech can provide information for designing experimental tasks (Demuth, 1998). The latter aspect becomes in particular relevant for the experiment on the comprehension of adjectives which is reported in Chapter 5.

This chapter is organized as follows: the research questions and the predictions following from the hypotheses presented in Chapter 2 and Chapter 3 are repeated in Section 4.1.

[1] Parts of this study have been published in Weicker and Schulz (2019).

Section 4.2 gives some background information on the spontaneous speech corpus and describes the procedure of the analysis. Section 4.3 describes the composition of the corpus and examines the influence of semantic complexity and the influence of the input on the order of acquisition of adjectives in spontaneous speech. The results are discussed in Section 4.4.

4.1 Research questions

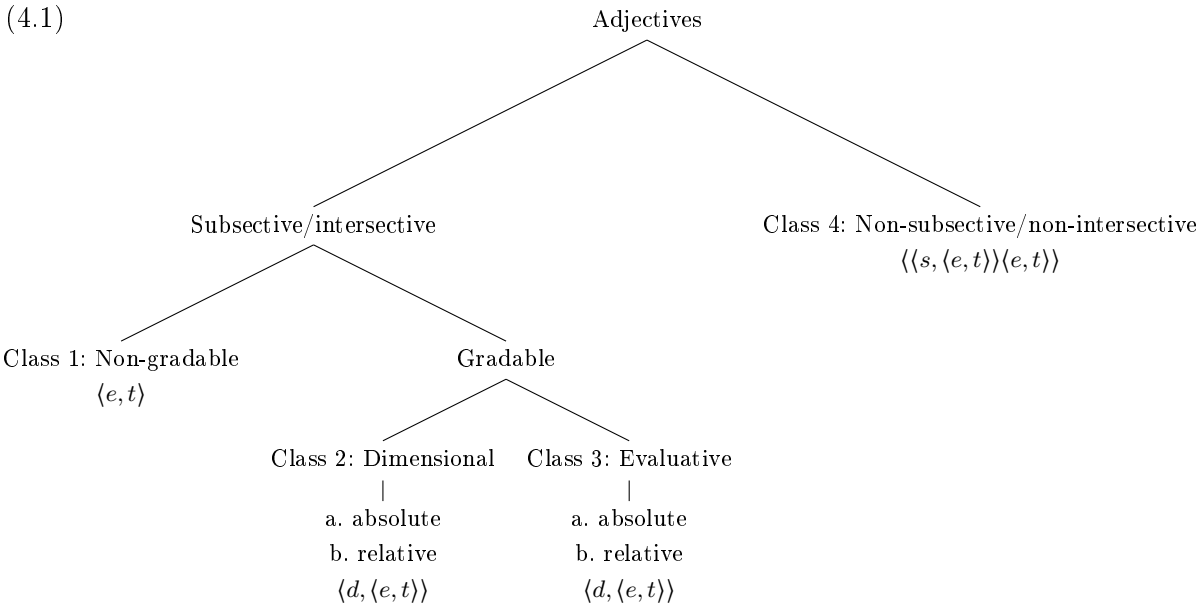
Hypothesis H1 introduced in Chapter 2 states:

(H1) The order of acquisition of adjectives is determined by their semantic complexity.

As a consequence, more complex adjectives according to the *Semantic Complexity Hierarchy* developed in Chapter 2 should not be acquired before less complex adjectives according to the *Semantic Complexity Hierarchy*. Adjectives of the same complexity should be acquired simultaneously.

The semantic classes proposed in Chapter 2 are given in (4.1). Recall that Class 1 is argued to be semantically least complex and Class 4 is argued to be semantically most complex. The subclasses a. and b. do not differ in semantic complexity.

(4.1)



With respect to the emergence of adjectives in spontaneous speech the following research questions result from H1:

(P1) Is the *Semantic Complexity Hierarchy* reflected in the mean age of acquisition for the adjective classes?

(P2) Is the *Semantic Complexity Hierarchy* reflected in the growth patterns of the adjective classes?

Regarding (P1), H1 predicts that the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition for the adjective classes. In particular, Classes 2a ('CLEAN') and 2b ('BIG') should not differ in their mean age of acquisition because they are equally complex. The same outcome is predicted for Classes 3a ('HEALTHY') and 3b ('BEAUTIFUL'). Class 1 ('BLUE') is predicted to have a lower or the same mean age of acquisition than Class 2. Class 2 should have a lower or the same mean age of acquisition than Class 3, and Class 3 should have a lower or the same mean age of acquisition than Class 4 ('FORMER'). Recall that the classes are sometimes named like typical adjectives belonging to the respective classes for ease of understanding. These are written in small caps. Note that each adjective class is assumed to comprise all adjectives that share the same specification of properties.

Growth in (P2) is understood as the age at which the number of types for each class increased most – a definition taken from the notion of vocabulary spurt or vocabulary explosion in lexical acquisition (e.g., E. V. Clark, 1993; Menyuk et al., 1995). This phenomenon describes that children acquire a large amount of words in a short period of time. It is likely that this phenomenon also appears for specific word classes, in this case adjectives. H1 predicts that due to their identical semantic complexity the "explosion" should occur at the same age for Classes 2a ('CLEAN') and 2b ('BIG') and for Classes 3a ('HEALTHY') and 3b ('BEAUTIFUL'), respectively. As a consequence of the differences in semantic complexity, Class 1 ('BLUE') should "explode" earlier than or simultaneously with Class 2. Class 2 should "explode" earlier than or simultaneously with Class 3. Class 3 should should "explode" earlier than or simultaneously with Class 4 ('FORMER').

Previous studies considered the input as a factor that determines the acqusition order of adjectives. Although it is likely that the input plays a role in the acquisition of the mental lexicon, and hence for the acquisition of adjectives, I assume that it is not the major factor that determines the acquisition order of adjectives as stated in H2.

(H2) The order of acquisition of adjectives is not completely determined by the input.

This hypothesis concerns the effect of the input frequency on children's production of adjectives in spontaneous speech. The following research questions are addressed:

(P3) Is adjective frequency in child and parental speech correlated?

(P4) Is adjective frequency in parental speech reflected in the adjectives' age of acquisition?

(P5) Are adjectives that children acquire late missing in parental speech before the age of acquisition?

Research question (P3) addresses the general influence of the adjective use in the input on the child's adjective use. According to hypothesis H2, it is possible that adjective frequency in child and parental speech are positively correlated suggesting a relation between child and parental adjective use. On the other hand, it is also possible that type and token frequency are distributed differently across adjective classes and child and parental

speech. Different distributions would indicate an independent adjective use in child and parental speech. Research question (P4) addresses the influence of the adjective use in the input on the acquisition order of the semantic classes proposed in Chapter 2. According to hypothesis H2, it is possible that the more frequent an adjective occurs in parental speech the earlier it is acquired by the child, but this negative correlation is not necessarily predicted by H2. Moreover, it may also be that frequency and semantic complexity coincide (see Brown & Hanlon, 1970, on syntactic complexity and input frequency). In this case, frequency does not need to determine the acquisition order. Hypothesis H2 further predicts that parents do not necessarily adjust their adjective use to the child's knowledge. Hence, it is possible that parents produce adjectives although these adjectives are not yet acquired by the child (P5).

4.2 Method

4.2.1 Data

This production study investigates the use of adjectives in the spontaneous speech of one monolingual German-speaking child ('Leo') (Behrens, 2006). The Leo-corpus contains transcripts of audio recordings and additional diary notes taken by the child's caregivers. Leo's parents have a higher education background and speak dialect-free standard High German (Behrens, 2006). The transcripts of the recorded child-caregiver interactions are available at the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). The corpus consists of almost half a million word forms produced by the child and three times as many word forms produced by his caregivers. The transcripts are morphosyntactically annotated, for instance, for parts-of-speech.

For the present study, the production of adjectives between the age of 2;00 and 2;11 years was analyzed. This age range was chosen because previous research indicated that children start producing adjectives around age two and that they acquire adjectives at a high pace during their third year of life (Tribushinina, Gillis, & De Maeyer, 2013). Furthermore, findings by Tribushinina, Gillis, and De Maeyer (2013) suggest that the syntactic complexity of structures in which adjectives occur increases from adjectives used in single-word utterances to adjectives used in target-like structures. The audio recordings in the Leo-corpus started when Leo was about to produce two-word utterances; between 2;00 and 2;11 his MLU increased steadily (Behrens, 2006).

During the time period investigated spontaneous speech samples of 60 minutes each were recorded five times per week in different home settings. These frequent and long recordings together with the caregiver's diary notes result in a very dense data set.

4.2.2 Procedure

The morphological coding and the part-of-speech tagging of the transcripts in the Leo-corpus allowed to search for all words belonging to the word class ADJECTIVE. First, all

words tagged as 'adjective' were extracted by using the CLAN command *freq*.[2] A frequency list including 33,119 adjective tokens[3] distributed over 559 adjective types[4] was generated.

If a word from the frequency list was not an open-class word or a content word, e.g., a determiner, possessive pronoun or cardinal number, it was excluded from further analysis. Ordinal numbers were included in the analysis because they behave semantically and morphologically similar to other prenominal modifiers (e.g., Wiese, 2003). Participles were excluded from the analysis because they show both adjectival and verbal properties. Like adjectives, participles are inflected in prenominal position. However, participles – like verbs – can be modified by manner adverbials implying that they refer to events (Rapp, 2015).

Second, for the adjectives from the frequency list that satisfied the inclusion criteria it was determined how they were used in the utterance and in the discourse. The CLAN command *kwal*[5] delivered the child's utterance in which the adjective of interest occurred and some lines before and after the target utterance. This additional context was helpful "to clarify the meaning of unclear utterances and to identify imitations, repetitions [...]" (Stromswold, 1998, p. 29). For all analyses, imitations exemplified in (4.2), repetitions exemplified in (4.3), formulaic expressions (e.g. *gute Nacht* 'good night'), and quotations from songs or poems were excluded (see Bittner, 2002; Nelson, 1976; Tribushinina et al., 2014).

(4.2) Leo, 2;08
*VAT: na, so eine kolossale Brücke, hä@o?
'*VAT: such a huge bridge, huh?'

*CHI: so ein-e kolossal-e Brücke-ø.
such a-NOM.SG.FEM huge-NOM.SG.FEM bridge-NOM.SG.FEM

'*CHI: such a huge bridge.'

(4.3) Leo, 2;02

*CHI: groß-er VW-Käfer-ø.
big-NOM.SG.MASC VW-Beetle-NOM.SG.MASC

'*CHI: big VW-Beetle.'
*CHI: großer VW-Käfer.

Moreover, all analyses were restricted to adjectives that specified the entities the speaker referred to or that assigned a property to an entity, hence that referred to a noun phrase.

[2] Version "freq (13-Jan-2015)".
[3] Tokens are defined as individual occurrences of a lexical unit and refer to the total number of adjectives in the corpus (Bußmann, 2008).
[4] Types are defined as lexems, i.e., as abstract units underlying different word forms (Bußmann, 2008). For example, the word forms *grün-e, grün-es, grün-er, grün* are all realizations of the same type *grün* ('green').
[5] Version "kwal (13-Jan-2015)"

Therefore, adverbially used adjectives as in (4.4) were not taken into account because they refer to events denoted by a verb phrase. Adjectives in (resultative) secondary predications as in (4.5) were also excluded because they do not describe a property of an entity directly, but the property is the result of an event described by the main predicate, hence the verb phrase is relevant here, too.

(4.4) Leo, 2;09

*CHI: wenn (e)s brenn-t komm-t (sch)nell die
when it.NOM.SG.NEUT burn-3SG come-3SG quick the.NOM.SG.FEM
Feuerwehr-ø.
fire brigade-NOM.SG.FEM

'*CHI: in case of fire the fire brigade arrives quickly .'

(4.5) Leo, 2;10

*CHI: rot-ø, die Haifischflosse-ø hab-(e) ich blau-ø
red-ø the.ACC.SG.FEM shark fin-ACC.SG.FEM have-1SG I.NOM.SG blue-ø
ge-mal-t.
PTCP-paint-PTCP

'*CHI: red, I painted the shark fin blue.'

Overall, 11,138 adjective tokens distributed over 323 adjective types remained. A complete list of the adjectives produced is given in Appendix A.1.

4.3 Analyses

In what follows, Section 4.3.1 gives an overview about Leo's adjective use at three different levels. At the conceptual level, I make use of notional adjective classes, primarily to compare the findings for the present corpus to findings from previous studies on the development of adjectives in spontaneous speech (Blackwell, 2005; Tribushinina et al., 2014). At the syntactic level, I examine in which syntactic structures Leo produced adjectives and how syntactic structures containing adjectives developed across age. At the semantic level, I transfer the notional classes to the semantic classes proposed in Chapter 2 to approach the question which compositional and lexical semantic properties underlie the concepts that constitute the lexicon.

Section 4.3.2 addresses research questions (P1) and (P2) and tests the hypothesis (H1) that the acquisition order of adjectives is influenced by their semantic complexity. Section 4.3.3, adresses research questions (P3) to (P5) and relates to the hypothesis (H2) that the order of acquisition of adjectives is not completely determined by their frequency in the input.

As spontaneous speech data from one child were analyzed, the corpus contains missing data points because there are adjectives that were not attested in every transcript. Hence,

the data set restricts the possibilities to compute inferential statistics. Therefore, some analyses remain descriptive. Where inferential statistics was employed, non-parametric tests were applied: for the comparison of two groups Mann-Whitney-U tests were used for independent samples, Wilcoxon tests were used for dependent samples. Kruskal-Wallis tests were used for the comparison of more than two independent samples. Frequency distributions for categorial variables were tested via Pearson Chi^2 tests. To test the relationship between two variables Spearman correlations were used.

4.3.1 Description of the corpus

As mentioned above, 11,138 adjective tokens were included in the analysis of Leo's spontaneous speech between 2;00 and 2;11 years. During this period Leo produced 296,589 tokens in total. Hence, the proportion of adjectives in Leos productive vocabulary is 3.8%. This value is similar to findings by Kauschke (2000) for German-speaking children at a similar age.

4.3.1.1 Conceptual level

Previous studies investigating the acquisition of adjectives in spontaneous speech focused on notional adjective classes (Blackwell, 2005; Tribushinina et al., 2014). As a result, the findings concern the acquisition of the concepts which adjectives express. Blackwell (2005) and Tribushinina et al. (2014) found a similar order of emergence of notional classes in the languages they studied (English, Dutch, German, French, Hebrew, Turkish). PHYSICAL PROPERTY, COLOR, SPATIAL PROPERTY, EVALUATIVE and CONFORMITY adjectives were among the early occuring classes. They emerged before INTERNAL STATE, BEHAVIOR, TEMPORAL and MODAL adjectives as well as before ordinal numbers.

The adjectives produced by Leo were classified with respect to the notional classes used in Blackwell (2005) and Tribushinina et al. (2014). The classes NATIONALITY and POSITION adjectives were added for the present data. Table 4.1 summarizes the classes used for the notional coding and gives examples from Leo's speech. The German equivalents are given in Appendix A.1.

Table 4.1: Adjective classes in the notional coding of adjectives based on Blackwell (2005) and Tribushinina et al. (2014).

Notional adjective class	Examples
AGE	*young, old*
BEHAVIOR	*wild, shy, friendly, funny*
COLOR	*blue, green, colorful*
CONFORMITY	*other, same, different, normal, correct*
DIMENSION	*big, thin, heavy, narrow*
MENTAL STATE	*sad, clever, disappointed, pleased*
NATIONALITY	*German, French, Hessian*
ORDINAL NUMBERS	*first, second, sixth*
PHYSICAL PROPERTY	*loud, fast, soft, dirty, dry, round, hot, sweet, fluffy*
PHYSICAL STATE	*healthy, hungry, dead, blind*
POSITION	*front, back*
QUANTITY	*alone, only, half*
TEMPORAL PROPERTY	*last, next*
VALUE	*good, beautiful, tasty, interesting*

Results

As illustrated in Table 4.2 Leo produced 10,692 adjective tokens distributed over 249 adjective types that were classified according to the notional classes in Table 4.1. The 74 types that were excluded from the analysis did either not fit into one of the classes in Table 4.1 (e.g., *teuer* 'expensive'), or they were not monomorphematic (e.g., *auszieh-bar* 'extensible'), or they were neologisms created by the child (e.g., *brüllerig*).

The observed frequencies for the notational classes are displayed in Table 4.2. Regarding the number of types, PHYSICAL PROPERTY adjectives constituted the largest notional class with 75 types. TEMPORAL PROPERTY adjectives were the smallest class including 2 types. The number of tokens was distributed differently across notional classes, too ($\chi^2(13, N = 249) = 43.206$, $p < .001$.). Post-hoc tests[6] revealed a significant difference between the number of tokens for the classes BEHAVIOR and PHYSICAL PROPERTY ($z = 3.692$, adjusted $p = .023$), BEHAVIOR and DIMENSION ($z = 3.498$, adjusted $p = .049$), BEHAVIOR and COLOR ($z = 5.081$, adjusted $p < .001$), MENTAL STATE and COLOR ($z = 4.123$, adjusted $p = .004$), ORDINAL NUMBER and COLOR ($z = 3.525$, adjusted $p = .044$). Descriptively, COLOR and DIMENSION adjectives were produced most frequently. POSITION and BEHAVIOR adjectives occurred very rarely. The distribution of types and tokens across notional classes in the present corpus is similar to previous findings by Blackwell (2005) and Tribushinina et al. (2014) as described in Chapter 3. The results on the development of tokens and types across age is given in Appendix A.2 and A.3. The number of types, but not the percentage of tokens, increased with age for all notional classes .

[6] Dunn-Bonferroni tests were used for pairwise post-hoc comparisons.

Table 4.2: Leo: Distribution of types and tokens across notional adjective classes.

Adjective class	Types		Tokens	
	Raw number	%	Raw number	%
AGE	4	1.6	700	6.5
BEHAVIOR	22	8.8	79	0.7
COLOR	16	6.4	2,823	26.4
CONFORMITY	12	4.8	1,249	11.7
DIMENSION	20	8.0	2,624	24.5
MENTAL STATE	19	7.6	153	1.4
NATIONALITY	11	4.4	112	1.0
ORDINAL NUMBERS	12	4.8	97	0.9
PHYSICAL PROPERTY	75	30.1	1,929	18.0
PHYSICAL STATE	14	5.6	281	2.6
POSITION	8	3.2	41	0.4
QUANTITY	11	4.4	117	1.1
TEMPORAL PROPERTY	2	0.8	123	1.2
VALUE	23	9.2	364	3.4
Total	249	100	10,692	100

When comparing the number of types and the number of tokens, it is evident that they are not evenly distributed. The class comprising the most types (PHYSICAL PROPERTY) is not the class with the most tokens and the class with the fewest types (TEMPORAL PROPERTY) is not the class with the fewest tokens. Some classes, e.g., VALUE, comprise a larger number of types compared to other classes, but fewer tokens. Other classes, e.g., AGE, consists of only few types, but are used frequently compared to other classes.

The acquisition order of the notional classes was determined based on the mean age of acquisition for each class (for details see Appendix A.4). The age of acquistion for each individual adjective was defined as the age of repeated use (see Stromswold, 1998), i.e., when the respective adjective was attested five times overall or two times within one month. Out of the 249 adjectives 182 satisfied the repeated use criterion. The age of acquisition was affected by the notional class ($\chi^2(13, N = 182) = 39.174$, $p < .001$), resulting in the following descriptive order of acquisition:

(4.6)
1. COLOR
2. DIMENSION
3. TEMPORAL PROPERTY
4. PHYSICAL PROPERTY, AGE
5. VALUE
6. CONFORMITY, PHYSICAL STATE, MENTAL STATE
7. NATIONALITY
8. POSITION, BEHAVIOR

9. QUANTITY
10. ORDINAL NUMBERS

Post-hoc tests revealed that the age of acquisition differed significantly between COLOR adjectives and ORDINAL NUMBERS (z = -3.569, adjusted p = .033) as well as between COLOR and QUANTITY adjectives (z = -3.543, adjusted p = .036). The order of acquisition found in the Leo-corpus is similar to the general acquisition path in the languages studied by Blackwell (2005) and Tribushinina et al. (2014). Two minor exceptions are TEMPORAL PROPERTY adjectives, which were on average acquired earlier by Leo, and CONFORMITY adjectives, which Leo acquired later than the children investigated in previous studies. In summary, it can be stated that the productive lexicon consists of the same property concepts and develops similarly in the third year of life across learners and across learners of different languages.

As mentioned above, it is open whether the interaction of adjective classes and the syntactic position they favor influence the order of acquisition. Previous studies on notional adjective classes analyzed either the interaction between adjective classes and syntactic position or the order of acquisition by investigating the global use of adjectives. A prerequisite for taking both aspects into account is to analyze the distribution of adjectives across syntactic contexts and the development of adjective structures with age.

4.3.1.2 Syntactic level

The aim of the present syntactic analysis is to investigate in which syntactic structures adjectives are used, how they develop with age, and whether an interaction between notional class and syntactic position can be found in German (see Blackwell, 2001, for English). Given the findings of previous studies with two-year-old children (Becker, 2000; Bittner, 1998; Clahsen et al., 1994), I expect to find DPs with two prenominal elements, e.g., with a determiner and an adjective, and utterances with copular verbs and adjectives. However, it is open when they emerge and how they are distributed because previous findings are inconclusive (Nelson, 1976; Blackwell, 2001).

For the syntactic analyses a total of 8,539 utterances satisfied the inclusion criteria described in the following. Only adjectives that can occur in both attributive and predicative position were included in the analysis. Hence, for instance *letzt-* ('last') or *nächst-* ('next') were excluded because they can appear only in attributive position. Adjectives such as *bereit* ('ready') or *allein(e)* ('alone') were excluded because they are possible only in predicative position. These adjectives would skew the picture. For instance, if an adjective that can only appear in attributive position is produced frequently, then the total number of attributive uses would also increase.

Occurrences with ambiguous syntactic structures were excluded, too. In the utterance in (4.7) *kleine* was counted as attributive, but for *große* it could not be determined whether it was a single-word utterance or whether the child corrected himself as indicated by [//] and a noun could possibly have followed the adjective. In the conversations in (4.8) - (4.14),

the child could have reacted to the preceding utterance with a more complex structure. However, all utterances (except for example (4.11)) are pragmatically felicitous, thus it is likely that the child uttered the less complex syntactic structure due to pragmatic reasons rather than due to the inability of producing complex syntactic structures.

(4.7) Leo, 2;02

*CHI: Groß-e [//] klein-e Brezel-ø.
big-? small-NOM.SG.FEM pretzel-NOM.SG.FEM

'*CHI: big [//] small pretzel.'

(4.8) Leo, 2;00
MUT: Das ist die +...
'*MUT: This is the ...'

*CHI: ++ groß-e.
big-NOM.SG.FEM

'*CHI: big one.'

(4.9) Leo, 2;01
MUT: Was für eine Farbe hat deine Hose?
'*MUT: Which color have your trousers?'

*CHI: grün-ø.
green-ø

'*CHI: green.'

(4.10) Leo, 2;00
MUT: Was ist das?
'*MUT: What is it?'

*CHI: heiß-ø.
hot-ø

'*CHI: hot.'

(4.11) Leo, 2;03
MUT: Was ist mit dem Rad passiert?
'*MUT: What happened with the wheel?'

*CHI: kaputt-ø.
broken-ø

'*CHI: broken.'

(4.12) Leo, 2;02
MUT: Was sind das da für Sterne hier drauf?
'*MUT: What kind of stars are on it?'

*CHI: grün-e.
green-NOM.PL.MASC

'*CHI: green ones.'

(4.13) Leo, 2;05
MUT: ja, die kommen auf alle Fälle aus Griechenland mit dem ICE.
'*MUT: They come definetely with an ICE from Greece.'

*CHI: mit (ei)n-ø rot-en.
with a-ø red-?.SG.MASC

'*CHI: with a red one.'

(4.14) Leo, 2;09
MEC: Müssen wir (ei)ne kleine Rakete bauen.
'*MEC: We have to build a small rocket.'

*CHI: ein-e groß-e.
a-ACC.SG.FEM big-ACC.SG.FEM

'*CHI: a big one.'

Structures with unintelligible material (*xxx*) were also excluded from the analysis if the position could have been filled with a noun or a copula.

(4.15) Leo, 2;07

*CHI: xxx einfach zu klein-ø die Fleischküchle-ø.
xxx simply too small-ø the.NOM.PL.NEUT hamburgerNOM.PL.NEUT

'*CHI: The hamburgers xxx simply too small.'

(4.16) Leo, 2;06

*CHI: (ei)n-ø gelb-es xxx.
a-NOM.SG.NEUT yellow-NOM.SG.NEUT xxx

'*CHI: a yellow xxx.'

Results

The 8,539 utterances were coded for the existence of a noun or pronoun, the existence of a functional element (definite and indefinite determiner, possessive pronoun, quantifier), and the existence of a copular verb. In addition, the position of the adjective relative to the noun was coded. It was not considered whether the adjective was inflected or not and

if it was inflected whether it showed the correct agreement morphology. The utterances in (4.17) to (4.30) exemplify the 14 different surface positions of adjectives that were found in Leo's speech. The positions are labeled in small caps[7] in the following. All utterances that contain a noun and a prenominal adjective or that are clearly elliptical were subsumed as attributive. Examples for attributive uses are given in (4.19), (4.22) and (4.24). Utterances with a copular verb (*sein* 'be', *werden* 'become', *bleiben* 'remain') or a copular-like verb (*aussehen* 'look', *schmecken* 'taste', *riechen* 'smell') and evaluative statements with *finden* ('find') were coded as predicative (Bußmann, 2008) independent of whether a noun was present or not. Postnominal adjectives were coded as predicative if adjective and noun were separated by lexical material, e.g., a focus particle, which indicated that the adjective is not adnominal. Examples for predicative uses are given in (4.21), (4.29), and (4.30). In what follows, the 14 syntactic structures involving adjectives are presented in detail.

In the corpus, adjectives appeared in isolation without a noun (4.17), or together with an adverb or a focus particle (4.18), respectively.

(4.17) ADJ

a. Leo, 2;01

*CHI: groß-er.
big-NOM.SG.MASC

'*CHI: big one.'

b. Leo, 2;00

*CHI: blau-ø.
blue-ø

'*CHI: blue.'

(4.18) X ADJ/ADJ X

a. Leo, 2;01

*CHI: eng-ø hier.
narrow-ø here

'*CHI: narrow here.'

b. Leo, 2;01

*CHI: auch blau-ø.
also blue-ø

'*CHI: also blue.'

In contrast to previous studies (Nelson, 1976), adjectives in single-word utterances as in (4.17) were not coded as predicative for three reasons. First, in German constructions with adjectives and without a noun are possible such as elliptical structures in (4.24) or nominalizations such as *die Rot* ('the Red') for a red pencil. In this regard German differs

[7] These terms are also used in the result section (see Figure 4.1).

from English, hence the coding applied in Nelson (1976) is not appropriate for German. Second, the copula has a particular semantic function as shown in Chapter 2. The copula needs to be realized to establish the predicative relation between subject and adjective. Third, if it is assumed that predicative structures, unlike attributive structures, involve a functional projection PredP as shown in Chapter 2, one can only tell that this projection is present if it is overtly realized by the copula. The second and third reason are also relevant for the analysis of uninflected postnominal adjectives as in (4.20a).

Moreover, adjectives occurred together with a noun either in prenominal (4.19) or in postnominal position as in (4.20) and (4.21). Although the determiner is missing in the utterances in (4.19), this structure can be target-like in German because the determiner is not obligatory for plural nouns as in (4.19a) or mass nouns (e.g., *water*).

(4.19) ADJ N

a. Leo, 2;02

*CHI: ICE blau-e Tür-en.
ICE blue-ACC.PL.FEM door-ACC.PL.FEM

'*CHI: ICE blue doors.'

b. Leo, 2;03

*CHI: groß-er VW-Käfer-ø.
big-NOM.SG.MASC VW-beetle-NOM.SG.MASC

'*CHI: big VW-beetle.'

c. Leo, 2;03

*CHI: weiß-ø Brezel-n.
white-ø pretzel-?.PL.FEM

'*CHI: white pretzels.'

(4.20) N ADJ

a. Leo, 2;01

*CHI: Zug-ø grün-ø.
train-?.SG.MASC green-ø

'*CHI: train green.'

b. Leo, 2;01

*CHI: Brücke-ø grün-e
bridge-?.SG.FEM green-?.SG.FEM

'*CHI: bridge green.'

(4.21) N X ADJ

a. Leo, 2;00

*CHI: Baby auch gelb-ø.
baby-?.SG.NEUT also yellow-ø

'*CHI: baby also yellow.'

b. Leo, 2;02

*CHI: Leo total müde-ø.
Leo-NOM.SG.MASC total tired-ø

'*CHI: Leo utterly tired.'

In contrast to structures such as (4.21), postnominal adjectives as in (4.20) were not analyzed as predicative. These structures are not necessarily predicative because the parameter setting for the German adjective position may not be in place yet. In German, prenominal adjectives agree with the noun, whereas predicative adjectives do not. Because both inflected and uninflected adjectives occurred postominally, it cannot be conluded that postnominal adjectives are predicative forms with the copula missing. Out of the 49 adjective types that were produced in postnominal position, 11 adjective types occurred inflected as well as uninflected as in the following examples: *Ampel rot* ('traffic light red') vs. *VW-Käfer rot-er* ('VW-beetle red'), *ein ICE rot* ('an ICE red') vs. *Luftballon rot-er* ('balloon red'), *Bagger groß* ('digger big') vs. *Herz groß-es* ('heart big'), *VW-Käfer klein* ('VW-beetle small') vs. *Tram klein-e* ('tramway small'). Hence, postnominal adjectives were not restricted to specific adjectives or nouns and should not be analyzed as frozen forms.

Adjectives in prenominal position (and very rarely in postnominal position) also occurred together with a functional element (e.g., an indefinite determiner) as shown in (4.22). These structures are an example for a target-like structure in German: the adjective is part of the DP and agrees with the noun in gender, number, and case. As mentioned before, agreement errors as in (4.22a) were not coded.

(4.22) DET ADJ N

a. Leo, 2;03

*CHI: ein-ø grün-ø Wagen-ø.
a-NOM.SG.MASC green-ø wagon-NOM.SG.MASC

'*CHI: a green wagon.'

b. Leo, 2;05

*CHI: eine Babykuh, ein-ø wild-er Stier-ø.
a baby cow a-NOM.SG.MASC wild-NOM.SG.MASC bull-NOM.SG.MASC

'*CHI: a baby cow, a wild bull.'

Adjectives also occurred together with a determiner but without a noun (4.23). This construction was treated separately from elliptic structures like the one in (4.24).

(4.23) Det Adj
Leo, 2;03
*MEC: das is(t) ganz schön doll nass, sonst hätten wir das noch nicht wechseln brauchen.
'this one is really wet, otherwise it would not have been necessary to change it.'

*CHI: ein-ø trocken-es.
a-NOM.SG.NEU dry-NOM.SG.NEU

'*CHI: a dry one.'

(4.24) Det Adj [N]
Leo, 2;09

*CHI: ein-en Gabelstaplerø, Billi hat auch
a-ACC.SG.MASC forklift-ACC.SG.MASC Billi-NOM.SG.MASC have-3SG also
ein-en groß-en.
a-ACC.SG.MASC big-ACC.SG.MASC

'*CHI: a forklift, Billi has also a big one.'

A structure was coded as elliptic if the noun could be reconstructed from the preceding discourse as it is the case in (4.24), but not in (4.23). Elliptic utterances are a licit structure in German.

There were also utterances in which lexical material was repeated. This can be the adjective as in (4.25) or the noun as in (4.26). These structures were analyzed separately because it cannot be determined whether the adjective is in pre- or in postnominal position.

(4.25) Adj N Adj
Leo, 2;02

*CHI: klein-e Weintraube-ø klein-e.
small-?.SG.FEM grape-?.SG.FEM small-?.SG.FEM

'*CHI: small grape small.'

(4.26) N Adj N
Leo, 2;01

*CHI: Löffel-ø groß-er Löffel-ø.
spoon-NOM.SG.MASC big-NOM.SG.MASC spoon-NOM.SG.MASC

'*CHI: spoon big spoon.'

Utterances in which the determiner was produced twice as in (4.27) or in which the adjective occurred before the DP as in (4.28) constituded their own category because it seems that adjective and noun are not intergrated and do not form a constituent.

(4.27) Det Adj Det N

a. Leo, 2;02

*CHI: ein-ø VW-Käfer-ø ein-ø
a-NOM.SG.MASC VW-beetle-NOM.SG.MASC a-NOM.SG.MASC
gelb-er.
yellow-NOM.SG.MASC

'*CHI: a VW-beetle a yellow.'

b. Leo, 2;09

*CHI: der gelb-e,
the.NOM.SG.MASC yellow-NOM.SG.MASC the.NOM.SG.MASC
der Kran-ø.
crane-NOM.SG.MASC

'*CHI: the yellow, the crane.'

(4.28) Adj Det N

a. Leo, 2;03

*CHI: ganz dreckig-ø (.) der Becher-ø.
completely dirty-ø the.NOM.SG.MASC cup-NOM.SG.MASC

'*CHI: completely dirty (.) the cup.'

b. Leo, 2;04

*CHI: Cargo-ø (.) ein-ø klein-er (.) der
Cargo-NOM.SG.MASC a-NOM.SG.MASC small-NOM.SG.MASC
fährt in Bahnhof rein.

'*CHI: Cargo (.) a small (.) it drives into the station.'

In utterances containing an adjective and a copular verb (*sein* 'be', *werden* 'become', *bleiben* 'remain') or a copular-like verb (*aussehen* 'look', *schmecken* 'taste', *riechen* 'smell') the noun could be realized as in (4.30) or missing as in (4.29). The same holds for evaluative statements with *finden* ('find').

(4.29) Copula Adj

Leo, 2;02

*CHI: is(t) groß-ø.
be-3SG big-ø

'*CHI: is big.'

(4.30) N copula Adj

Leo, 2;04

*CHI: der Hecht is(t) groß-ø.
the.NOM.SG.MASC pike-NOM.SG.MASC be-3SG big-ø

'*CHI: the pike is big.'

It is a target-like option in German to use an adjective as the complement of a copula as shown in (4.30). In addition to the structures in (4.17) to (4.30), adjectives appeared after an indefinite pronoun in constructions like *was blaues* ('something blue'). These constructions are also possible in German and constituted an individual category.

Figure 4.1 displays the proportion of the syntactic structures described in (4.17)-(4.30) relative to the total number of analyzed utterances at different ages. Syntactic structures that occurred less than 3% at the respective age are subsumed under the label OTHER. The other labels refer to the respective structures in the examples named by the terms in small caps.

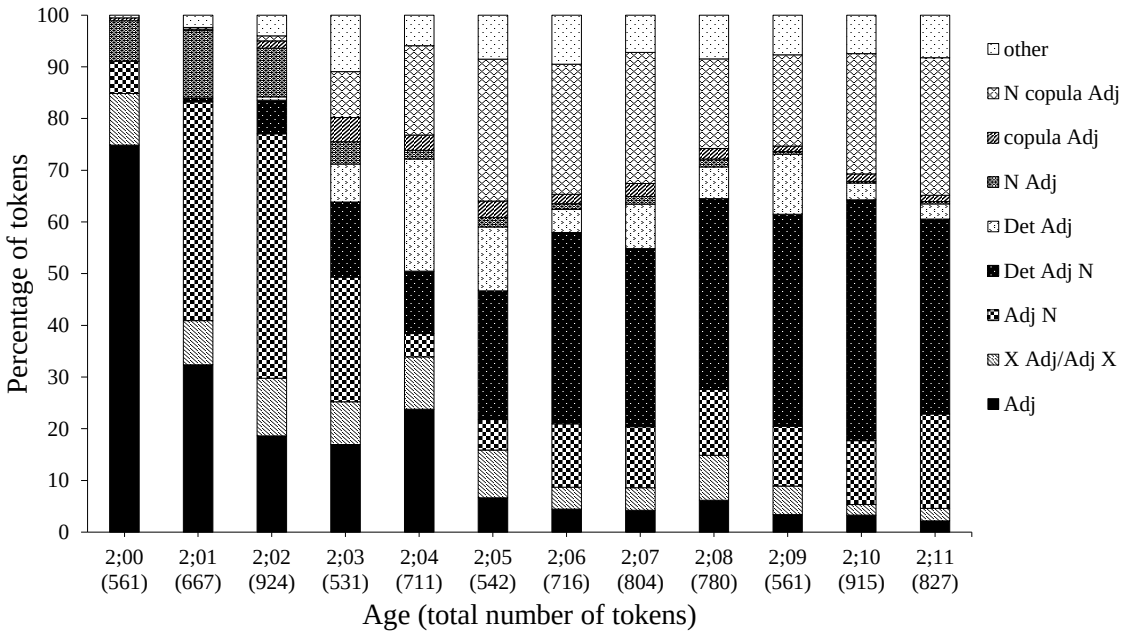


Figure 4.1: Leo: Distribution of adjectives across syntactic contexts.

Figure 4.1 shows that the number of adjectives in utterances lacking a noun (ADJ, X ADJ/ADJ X) decreased with age, whereas the number of adjectives in attributive (ADJ N, DET ADJ N) and predicative structures (COPULA ADJ, N COPULA ADJ) increased with age.

At 2;00, adjectives were mostly produced as single-word utterances (74.9%, $N = 420$). The proportion of single-word adjective utterances decreased with age to 2.2% ($N = 18$) at 2;11. For adjectives in isolation, the percentage of occurrences is negatively correlated with age (r_s = -.958, $p < .001$). In 14% of the utterances at 2:00, the adjective was combined with a noun: in 6.2% ($N = 35$) the adjective preceded the noun, in 7.8% ($N = 44$) the adjective followed the noun. In both structures a functional element such as a determiner is missing. The percentage of occurrences for prenominal adjectives is not correlated with age. This may be due to the fact that determinerless adjective-noun phrases are target-like in German if the phrase includes a mass noun or is indefinite. In contrast, the percentage

of occurrences for postnominal adjectives correlates negatively with age (r_s = -.937, $p <$ .001). At 2;11, 0.5% (N = 4) of the utterances consisted of a postnominal adjective. The decrease of postnominal adjectives can have two possible reasons: as argued above, it may be that the parameter setting was not in place, i.e., the child was in the process of learning that adjectives precede the noun in German. Alternatively, it may be that the child does not overtly realize copulas until a specific age. It has been found for Leo that the use of copulas increases at 2;04 (Behrens, 2006). However, because both inflected and uninflected adjectives occurred postnominally, this structure can interpreted both as a precusor of attributive and predicative structures.

The first prenominal adjectives preceded by a determiner were attested at 2;01 (0.7%, N = 5). The proportion of prenominal adjective structures containing a determiner is positively correlated with age (r_s = -.951, $p <$ .001). At 2;11, 37.7% (N = 312) of all utterances were DPs containing an adjective. This increase is consistent with previous findings for German (Bittner, 1998; Clahsen et al., 1994). The first predicative structures with a copula relating adjective and noun also emerged at 2;01 (0.3%, N = 2). Their proportion is also positively correlated with age (r_s = -.748, p = .005) and increased up to 26.6% (N = 220) until 2;11.

Adjectives as part of proper DPs as well as adjectives as complements of copulas emerged at 2;01. As described before, adjectives as part of DPs and elliptic utterances were analyzed as attributive. Adjectives as complements of copulas and postnominal adjectives with other lexical material between adjective and noun were analyzed as predicative. Out of the 8,539 utterances that were included in the syntactic analysis 5,724 utterances could be classified as attributive or predicative. 68.7% of these utterances were attributive structures and 31.3% were predicative structures. The uneven distribution of attributive and predicative structures also holds across age ($\chi^2(11) = 415.468$, $p <$ .001) as shown in Figure 4.2. The results show that it is not the case that one structure appeared earlier than the other. Note that in Figure 4.2 only the structures in examples (4.17) to (4.30) that could be classified as predicative or attributive were included. The dominance of attributive adjectives over predicative ones has also been observed for English (Blackwell, 2001). The reason why Nelson (1976) found the reverse distribution may result from a different coding with respect to single-word utterances.

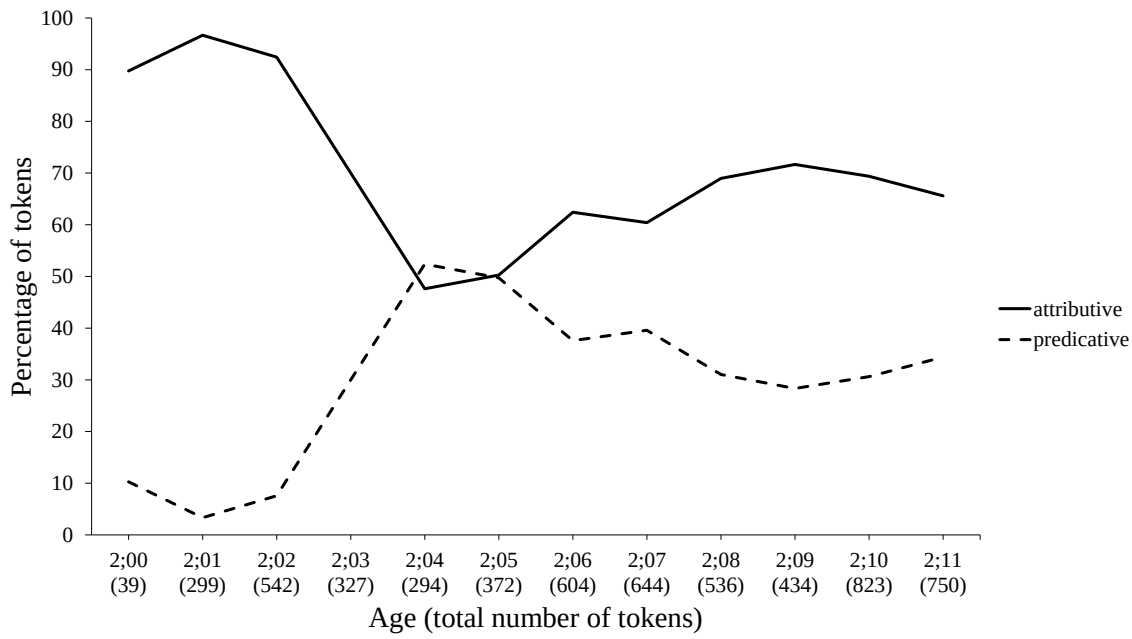


Figure 4.2: Leo: Percentage of attributive and predicative occurrences across age. 100% equates the sum of attributive and predicative structures at the respective age. Other adjective structures are excluded.

One exception to the distribution of attributive and predicative structures arised at 2;04. Up to this age, predicative structures increased. At the same time, attributive structures decreased with the result that for one month predicative structures were more frequent than attributive structures. As of age 2;05, attributive adjectives were again produced more frequently than predicative adjectives. Behrens (2006) describes that at 2;04, the use of copulas increases. This may explain the boost of predicative adjective structures, but leaves open why attributive adjective structures decline. One could speculate that at that age, Leo reorganized his system of modification structures in general and that he tried "new" structures. However, this conclusion is very tentative. More detailed investigations are necessary that cover the development of other modification structures (e.g., prepositional phrases or relative clauses) and the development of different types of noun phrases at this age range. As shown in Chapter 2, predicative adjectives are often used to comment on already established referents, so it may be interesting in this regard to investigate the acquisition of pronouns for instance.

Nelson (1976) and Blackwell (2001) found a relation between the syntactic position and the concept expressed by the adjective such that different notional classes are preferably used in different syntactic positions. Figure 4.3 illustrates that this relation between notional class and syntactic position also holds for the adjectives produced by Leo ($\chi^2(7) = 1548.192$, $p < .001$). This analysis focuses on the notional classes mentioned by Blackwell (2001) and takes all adjective types belonging to the respective classes into account (see Table 4.2 for the occurrences in Leo's data) independent of whether they are restricted to only one syntactic position. Note that the label OTHER in Figure 4.3 includes all structures

in examples (4.17) to (4.30) that could not be classified as predicative or attributive.

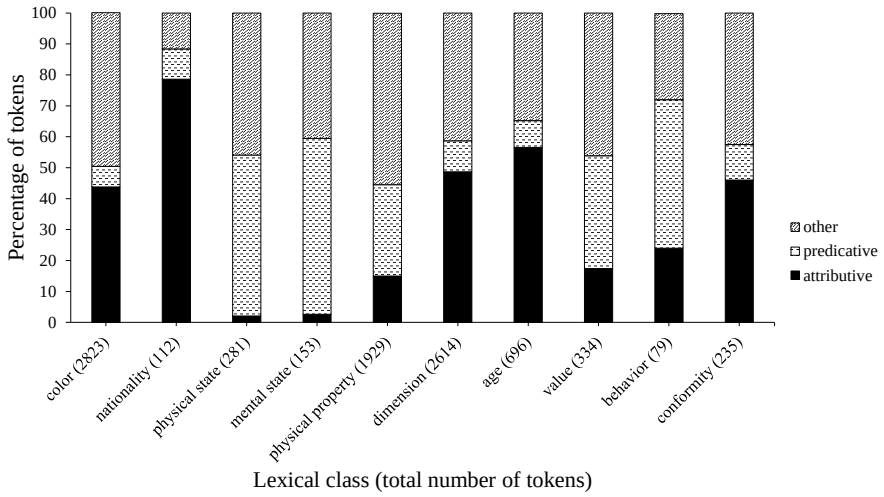


Figure 4.3: Leo: Percentage of attributive and predicative occurrences per notional class.

The following notional classes occurred more often in attributive than in predicative position: COLOR (43.8 vs. 6.6%), DIMENSION (48.7 vs. 6.6%), and AGE (56.6 vs. 8.6%). The difference between the number of attributive and predicative occurrences was statistically significant for COLOR (Z = -3.323, p = .001). A trend was observed for DIMENSION (Z = -1.948, p = .051). In contrast, the following notional classes occurred less often in attributive than in predicative position: PHYSICAL STATE (2.1 vs. 52.0%), MENTAL STATE (2.6 vs. 56.9%), VALUE (17.4 vs. 36.5%), BEHAVIOR (24.0 vs. 48.0%), and PHYSICAL PROPERTY (15.0 vs. 29.5%). The difference between the number of attributive and predicative occurrences was statistically significant for PHYSICAL PROPERTY (Z = -2.908, p = .004), PHYSICAL STATE (Z = -3.299, p = .001), and MENTAL STATE (Z = -3.464, p = .001). Except for VALUE adjectives, the same preferences were found in previous studies.

The findings regarding the relation between syntactic position and notional class may relate to the question why predicative adjectives are more frequent than attributive adjectives at 2;04. Besides structural reasons it is possible that at this age adjectives become more prominent that are preferred in predicative position. A case in point are PHYSICAL PROPERTY adjectives. As can be seen in Figure A.1 in Appendix A.2 their proportion increased between 2;01 and 2;04. In contrast, the proportion of DIMENSION adjectives, which were more often used attributively, decreased during that time. However, this relationship could also hold *vice versa*.

In summary, the syntactic analysis of Leo's utterances containing adjectives revealed that early adjectives were mainly produced in isolation, i.e., without a nominal element. Attibutive and predicative adjectives emerged at the same time, but attributive structures were preferred over predicative structures. However, this preference does not hold overall, but is related to the adjectives' notional properties. Section 4.3.1.3 discusses how the notional classes can be transferred into semantic classes that are based on entailment and

gradability properties of adjectives. Hence, the question is addressed which compositional and lexical semantic properties underlie the concepts that constitute Leo's productive adjective lexicon.

4.3.1.3 Semantic level

The aim of the present semantic analysis is twofold. First, it aims to provide a more comprehensive picture of the development of gradability and entailment properties than the studies that focused on selected aspects such as the acquisition of scalar structure (Tribushinina & Gillis, 2012) or the acquisition of comparison constructions (Hohaus et al., 2014). Tribushinina and Gillis (2012) found that non-gradable, absolute gradable, relative gradable and evaluative adjectives were attested in spontaneous speech from age 2. In contrast to the classification in this thesis (see Table 4.3), evaluative adjectives were not further distinguished regarding the scale type (open vs. closed). Because Tribushinina and Gillis (2012) used cross-sectional data, longitudinal developments could have been assessed only indirectly as noted by the authors. Hohaus et al. (2014) studied a related phenomenon: the acquisition of comparison constructions. In their study, gradable adjectives were not distinguished according to different scale types. Hohaus et al.'s findings suggest that the positive unmarked form of gradable adjectives is acquired between 2;00 and 2;05 years. The second aim of the present study is to relate the findings regarding the emergence of the more conceptual notional classes reported in Section 4.3.1.1 to the emergence of formal semantic classes.

The adjectives produced by Leo were coded according to both notional classes (see Section 4.3.1.1) and semantic classes. The semantic classes of interest are repeated in Table 4.3.

Table 4.3: Adjective classes in the semantic coding.

Adjective class	Semantic properties	Examples
Class 1	subsective, intersective, non-gradable	*blue, German, square, dead*
Class 2a	subsective, intersective, absolute gradable, dimensional	*clean, dry, broken, dotted*
Class 2b	subsective, intersective, relative gradable, dimensional	*big, expensive, weak, young*
Class 3a	subsective, intersective, absolute gradable, evaluative	*healthy, sad, happy, shy*
Class 3b	subsective, intersective, relative gradable, evaluative	*beautiful, good, industrious, nice*
Class 4	non-subsective, non-intersective	*former, alleged, possible, past*

In Chapter 2, I described criteria that can be used to diagnose whether an adjective belongs to one of the semantic classes. Note that the assignment of an adjective to one of

the classes was based on adult use. Hence, the underlying assumption was that an adjective in Leo's speech is produced with target-like semantics. The diagnostic criteria in Figure 2.2 are repeated in (4.31):

(4.31) D1. Does the adjective in attributive position describe a subset of the set denoted by the noun?

D2. Can the adjective occur in comparison constructions?

D3. Can the adjective be extended by *in jeder Hinsicht*?

D4. Can the adjective be modified by *vollständig* or *ein bisschen*?

D1 differentiates between subsective/intersective adjectives and non-subsective/non-intersective adjectives. If this question is negated, an adjective belongs to Class 4. Otherwise, the adjective can belong to Class 1, 2 or 3. D2 distinguishes between gradable and non-gradable subsective/intersective adjectives. If this question is negated (marked by a cross in Table 4.4), an adjective belongs to Class 1. Otherwise, the adjective belongs to Class 2 or 3. D3 divides gradable adjectives into dimensional and evaluative adjectives. Therefore, it is not applicable to Class 1 (marked as "na" in Table 4.4). D3 is based on the observation that evaluative adjectives are often multidimensional and can be modifed by phrases such as *in jeder Hinsicht* ('in every respect'). If this modification is not possible, an adjective belongs to Class 2; if it is possible, the adjective belongs to Class 3. D4 discriminates between absolute and relative gradable adjectives. As explained in Chapter 2, absolute gradable adjectives can be modified by *fast* ('almost') or *ein bisschen* ('slightly'), whereas relative gradable adjectives cannot. Hence, if modification by one of the degree words is not possible, an adjective belongs to class 2b or 3b depending on whether the third criterion applies or not. Otherwise, an adjective belongs to either Class 2a or 3a. Table 4.4 summarizes how the adjective classes behave with respect to the diagnostic criteria. Note that all adjectives produced by Leo fulfilled criterion D1, thus Class 4 is not present in his speech between 2;00 and 2;11. Therefore, in the remainder of the chapter Class 4 is not taken into account.[8] Whether an adjective fulfills the criteria in D2, D3, and D4 was tested via a Google search.[9]

[8] The first adjective from Class 4 is *frühere* ('former') attested at 3;04 while Leo plays on the floor with vehicles, bricks etc. saying: *das wird das frühere Florenz* ('this is going to be the Florence in former times').

[9] Lapata and Keller (2005) showed that web-based searches yield reliable results such that for instance web frequencies and corpus frequencies are highly correlated.

Table 4.4: Applicability of classification criteria D2-D4 to semantic classes 1-3.

Adjective class	D2: Occurrence in comparative/ superlative constructions	D3: Modification by *in jeder Hinsicht* ('in every respect')	D4: Modification by *vollständig*, *ein bisschen* ('completely', 'slightly')
1 'BLUE'	✗	na	na
2a 'CLEAN'	✓	✗	✓
2b 'BIG'	✓	✗	✗
3a 'HEALTHY'	✓	✓	✓
3b'BEAUTIFUL'	✓	✓	✗

Note. ✗= not possible; ✓= possible; na = non applicable.

Results

Out of the 323 adjective types produced by Leo 202 types could be assigned to one of the semantic classes. Table 4.5 illustrates how the types were distributed over the semantic classes and how frequent the specific classes were in Leo's speech.

Table 4.5: Leo: Distribution of types and tokens across semantic adjective classes.

Adjective class	Types		Tokens	
	Raw number	%	Raw number	%
1 'BLUE'	38	18.8	3096	33.3
2a 'CLEAN'	50	24.8	1312	14.1
2b 'BIG'	39	19.3	4017	43.2
3a 'HEALTHY'	50	24.8	471	5.1
3b 'BEAUTIFUL'	25	12.4	393	4.2
Total	202	100	9289	100

In terms of types, absolute gradable dimensional (2a 'CLEAN') and absolute gradable evaluative (3a 'HEALTHY') adjectives constituted the largest classes. The class of relative gradable evaluative adjectives (3b 'BEAUTIFUL') contained the fewest types. Out of the 202 adjective types 154 were used productively according to the criterion of repeated use. Figure 4.4 displays the development of productive types per semantic class. For each class the number of types is positively correlated with age (1: $r_s = .991$, $p < .001$, 2a: $r_s = .992$, $p < .001$, 2b: $r_s = .960$, $p < .001$, 3a: $r_s = .983$, $p < .001$, 3b: $r_s = .983$, $p < .001$).

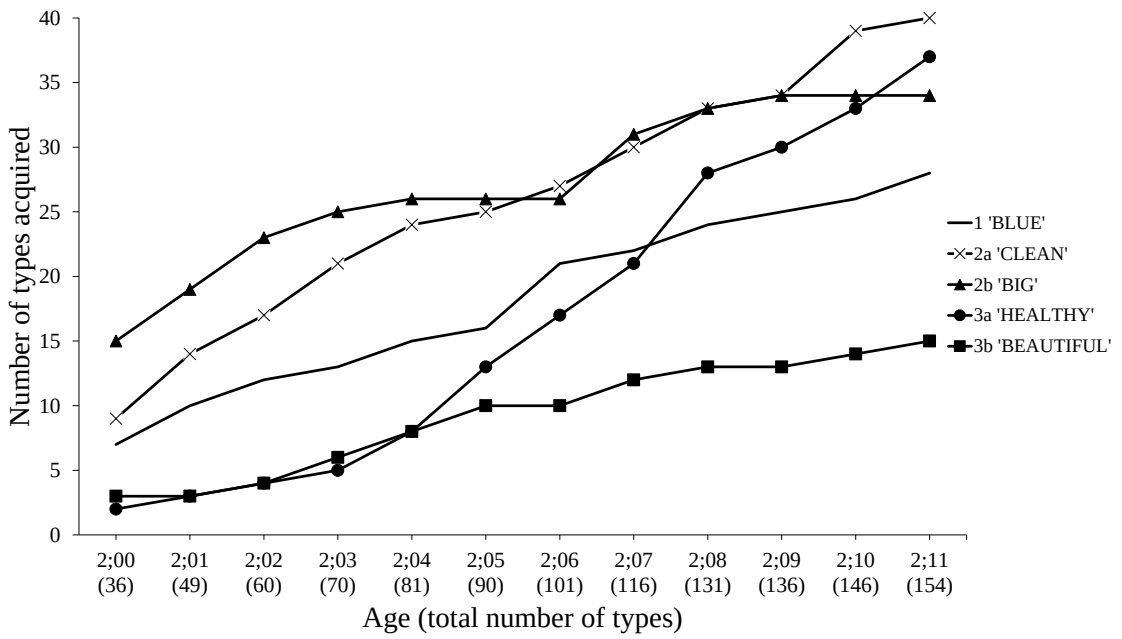


Figure 4.4: Leo: Distribution of adjectives types per semantic adjective class across age.

At 2;00, relative gradable dimensional adjectives (2b 'BIG') were the largest class. The number of types increased more steeply for absolute gradable dimensional (2a 'CLEAN') and evaluative (3a 'HEALTHY') adjectives than for relative gradable dimensional adjectives. They lagged behind at 2;11. The number of relative gradable evaluative adjectives (3b 'BEAUTIFUL') remained low throughout development.

The number of tokens was distributed differently across semantic classes ($\chi^2(4, N = 202) = 15.528, p = .004$): relative gradable dimensional adjectives (2b 'BIG') were produced most frequently, relative gradable evaluative adjectives (3b 'BEAUTIFUL') occurred least often. Post-hoc tests revealed that the total number of tokens differed significantly between relative gradable dimensional and relative gradable evaluative adjectives (z = 3.124, adjusted p = .018) as well as between relative gradable dimensional and absolute gradable evaluative adjectives (z = -3.234, adjusted p = .012). A closer look at the development over time shows that the overall number of adjective tokens varied across recordings, hence across age. Figure 4.5 displays the percentage of tokens per class relative to the total number of adjective tokens. The distribution of adjective tokens across age mirrors the overall distribution in Table 4.5: relative gradable dimensional adjectives (2b 'BIG') were produced most frequently followed by non-gradable adjectives (1 'BLUE') except for the latest time point. At 2;00, absolute gradable dimensional adjectives (2a 'CLEAN') were more frequent than non-gradable adjectives, but from 2;01 their relation was reversed. The proportion of absolute and relative gradable evaluative adjectives (3a 'HEALTHY' and 3b 'BEAUTIFUL') was similarly low. For none of the classes the percentage of tokens is correlated with age.

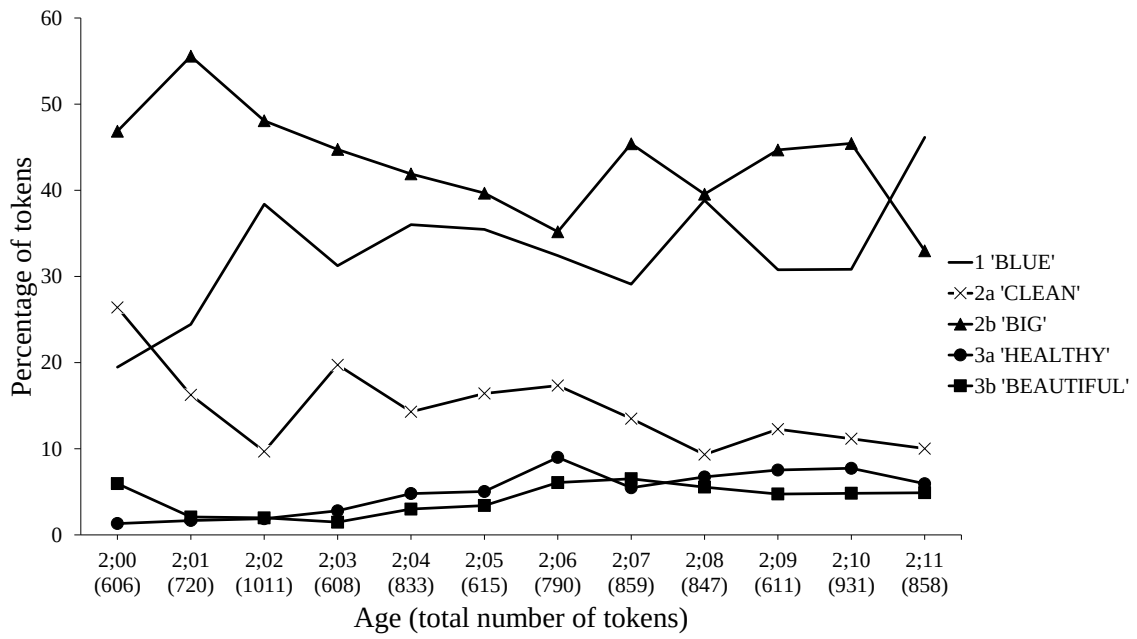


Figure 4.5: Leo: Distribution of adjective tokens per semantic adjective class across age.

The number of types and the number of tokens are not evenly distributed. The classes comprising the most types are absolute gradable dimensional and evaluative adjectives (2a 'CLEAN' and 3a 'HEALTHY') with 50 types each. However, these classes are not the classes with the most tokens. A class with a balanced distribution of types and tokens are relative gradable evaluative adjectives (3b 'BEAUTIFUL'): this class comprises the fewest types (N = 25) and the fewest tokens (N = 393).

Figure 4.3 (see Section 4.3.1.2) indicates that some notional classes are used preferably in attributive position, whereas other notional classes are used preferably in predicative position. The data presented in Figure 4.6 also suggest a relation between syntactic position and semantic class ($\chi^2(4) = 1021.273$, $p < .001$).

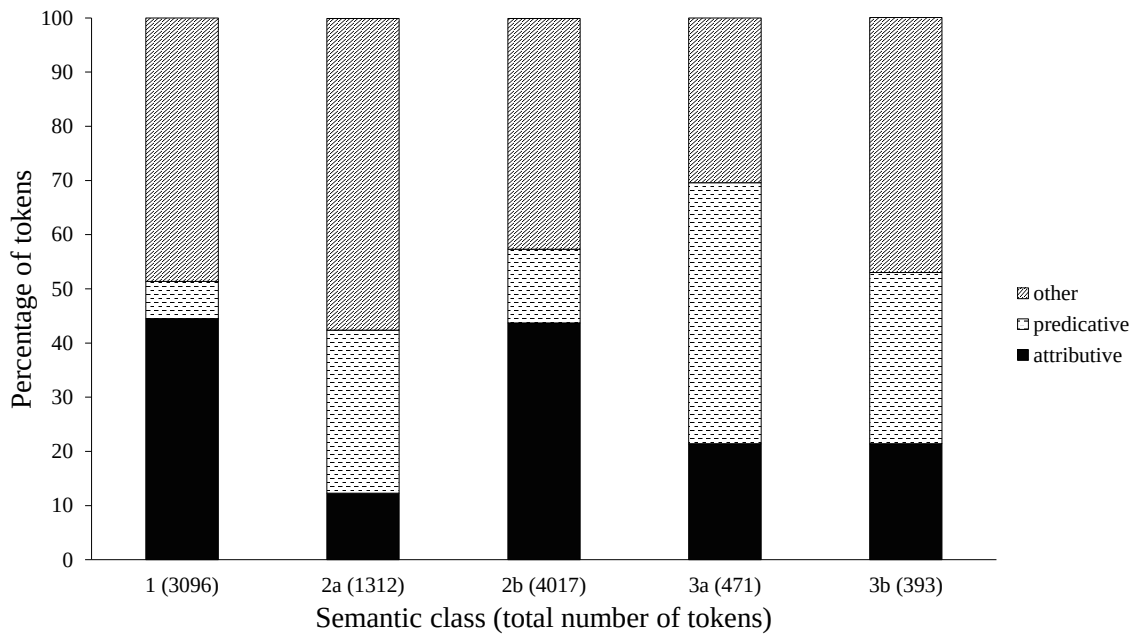


Figure 4.6: Leo: Percentage of attributive and predicative occurrences per semantic class.

The following semantic classes occurred more often in attributive than in predicative position: non-gradable (1 'BLUE') and relative gradable dimensional (2b 'BIG') adjectives (Class 1: 44.5 vs. 6.9%, Class 2b: 43.7 vs. 13.6%). The difference between the number of attributive and predicative occurrences was statistically significant for non-gradable adjectives (Z = -4.373 , $p < .001$). In contrast, the following notional classes occurred less often in attributive than in predicative position: absolute gradable dimensional (2a 'CLEAN') and evaluative (3a 'HEALTHY') (Class 2a : 12.3 vs. 30.2%, Class 3a: 21.4 vs. 48.2%). The difference between the number of attributive and predicative occurrences was statistically significant for absolute gradable dimensional (Z = -2.967 , p = .003), and absolute gradable evaluative adjectives (Z = -3.268, p = .001). A slight preference for the predicative use is observed for relative gradable evaluative adjectives (3b 'BEAUTIFUL') (21.4 vs. 31.6%).

The findings for notional and semantic classes suggest a connection between the two classifications. The distribution of adjective types and of tokens was different across notional and across semantic classes. A relation between syntactic position and notional class and a relation between syntactic position and semantic class was found in Leo's speech. However, how these classifications are related is not clear (see Chapter 2). It seems plausible to assume that notional classes express the concepts that compose the mental lexicon and that semantic classes describe the compositional and lexical meaning of these concepts. The semantic properties guide compositional operations and interpretive features, for instance whether an adjective must be interpreted relative to a standard of comparison. Out of all adjectives produced by Leo, 202 were coded for both semantic and notional class. Adjectives that were coded only for notional class were the ones which behave differently with respect to entailment and gradability properties, e.g., extreme adjectives such as *riesig*

('huge') or *wunderbar* ('marvellous') (see Morzycki, 2012), or ordinal numbers (see Bhatt, 1999; Bylinina, Ivlieva, Podobryaev, & Sudo, 2015; Wiese, 2003). For the adjectives that could be coded for both semantic and notional class it is possible to detect how they are related. The correspondence of semantic and notional classes that results from this double coding is illustrated in Table 4.6.

Table 4.6: Correspondence of semantic and notional adjective classes based on Leo-corpus.

Notional class	Semantic class					
	1	2a	2b	3a	3b	4
AGE			✓			
BEHAVIOR				✓	✓	
COLOR	✓					
CONFORMITY	✓			✓		
DIMENSION			✓			
MENTAL STATE				✓	✓	
NATIONALITY	✓					
PHYSICAL PROPERTY	✓	✓	✓	✓		
PHYSICAL STATE	✓			✓		
VALUE				✓	✓	

Note. ✓= Notional class corresponds to semantic class.

As predicted by Demonte (2011) (see Chapter 2), transferring notional classes into semantic classes does not yield a strict correpondence between the two classifications. PHYSICAL PROPERTY adjectives fall into various semantic classes. This is not surprising given the fact that this class is very large and heterogeneous. In contrast, COLOR and NATIONALITY adjectives can be clearly assigned to Class 1; DIMENSION and AGE adjectives fall exclusively into Class 2b. The classes MENTAL STATE, VALUE and BEHAVIOR correspond to Classes 3a and 3b, which differ only in the scale type.

The question is now whether the (partial) correspondence between notional and semantic classes is also mirrored in the findings regarding the relation between syntactic position and adjective class and regarding the overall distribution of types and tokens. The notional classes that were produced most frequently in terms of tokens were COLOR and DIMENSION (see Table 4.2). According to Table 4.6 they correspond to the semantic classes 1 and 2b. As can be seen in Table 4.5, these are the most frequent semantic classes with respect to the number of tokens. Hence, the findings partially match for notional and semantic classes. Regarding the number of types, PHYSICAL PROPERTY adjectives were the largest class followed by VALUE and BEHAVIOR adjectives. Because these classes fall into multiple semantic classes, the findings are difficult to compare. However, all three notional classes correspond to Class 2a or 3a, which are the largest semantic classes in terms of types. Thus, the findings match for notional and semantic classes although not as clear as for the distribution of tokens.

Regarding the relation between syntactic position and adjective class, the findings indicate that Leo produces the notional classes COLOR, DIMENSION and AGE more frequently in attributive position (see Figure 4.3). These notional classes correpsond to the semantic classes 1 and 2b, which are also more common in attributive position as illustrated in Figure 4.6. The notional classes PHYSICAL PROPERTY, PHYSICAL STATE, MENTAL STATE, VALUE and BEHAVIOR are mostly used predicatively matching the preference for predicative uses of the semantic classes to which they correspond (2a, 3a, 3b). But PHYSICAL PROPERTY and PHYSICAL STATE adjectives also fall in Class 1 and Class 2b – here, the findings for notional and semantic classes do not coincide.

4.3.1.4 Summary

One aim of this description of the Leo-corpus with regard to the adjective use between 2;00 and 2;11 years was to compare the findings for German with previous findings regarding the production of notional classes and regarding the syntactic structures in which adjectives are produced. More importantly, this part extended previous work on the emergence of adjectives in spontaneous speech by adding the acquisition of compositional and lexical semantic properties of adjectives. Moreover, the relationship between notional and compositional and lexical semantic properties of adjectives was addressed.

The analysis of notional classes revealed that the largest class in terms of types constituted PHYSICAL PROPERTY adjectives. COLOR and DIMENSION adjectives were most frequently produced in terms of tokens. Regarding the composition of the lexicon and the order of emergence of notional classes in spontaneous speech the present findings for German are consistent with findings from previous studies (Blackwell, 2005; Tribushinina et al., 2014), and hence suggest a similar acquisition pattern across learners and across learners of different languages.

The analysis of the syntactic stuctures in which Leo used adjectives revealed that the production of adjectives started with single-word utterances. When Leo began to combine adjectives and nouns, predicative and attributive structures emerged at the same age. Overall, attributive adjectives were attested more frequently than predicative adjectives between the age of 2;00 and 2;11 years. The use of attributive and predicative structures by notional class revealed different distributions across notional classes similar to previous findings from Blackwell (2001).

The analysis of semantic classes revealed that non-subsective/non-intersective adjectives were not produced. All other semantic classes were attested and increased with age in terms of types. Among the subsective/intersective adjectives, the largest class in terms of types constituted absolute gradable dimensional (Class 2a 'CLEAN') and absolute gradable evaluative (Class 3a 'HEALTHY') adjectives. Overall, relative gradable dimensional adjectives (Class 2b 'BIG') were most frequently produced in terms of tokens. Like notional classes, semantic classes were distributed differently across attributive and predicative uses.

The comparison of semantic and notional classes suggests a correpondence between

these two classifications. Notional classes have been suggested to express the concepts which constitute the adjective lexicon, whereas semantic classes refer to the compositional and lexical semantic properties underlying these concepts. Thus, semantic properties provide crucial information about the adjectives' behavior in combination with other lexical items, in particular nouns.

Recall that research questions (P1) and (P2) asked whether the *Semantic Complexity Hierachy* is reflected in the mean age of acquisition and in the growth patterns of the semantic classes. In the present part only type and token frequencies were analyzed. However, the two research questions do not imply that less complex adjectives are produced more frequently than more complex adjectives in the course of acquisition. The two research question rather focus on the order of acquisition. From a theoretical point of view (Hawkins, 1987), predictions about acquisition order do not entail this type of quantitative prediction. Second, regarding the number of types and tokens in one adjective class compared to another class, it may simply be that some adjectives are more relevant to express the child's communicative needs than others independent of the adjectives' complexity. The present findings indicate that semantic complexity is mirrored in neither type frequency nor token frequency. In addition, Blackwell (2005) found that child frequency was not a significant predictor of the age of acquisition, although early acquired adjectives were also produced more frequently by the children investigated.[10] Tribushinina et al. (2014) also note that adjective frequency, in their study operationalized as the probability of adjective use, is not the same as acquisition. Hence, to investigate whether the order of acquisition of semantic classes is influenced by the adjectives' semantic complexity as proposed by the *Semantic Complexity Hierarchy*, neither type nor token frequency alone are expected to yield reliable results. In Section 4.3.2, the order of acquisition is measured as the mean age of acquisition and in growth patterns. H1 predicts that these two measures mirror semantic complexity.

4.3.2 Semantic complexity and order of acquisition

Hypothesis H1 states that the order of adjective acquisition is determined by the adjectives' complexity. So far, the semantic classes in the *Semantic Complexity Hierarchy* were investigated regarding type and token frequencies only. As suggested in the previous section, frequency does not mirror complexity: semantic classes that are assumed to be less complex are not more frequent than semantic classes that are assumed to be more complex. For instance, relative gradable dimensional adjectives (Class 2b 'BIG') are more frequent in tokens than non-gradable adjectives (Class 1 'BLUE'), and absolute gradable dimensional adjectives (Class 2a 'CLEAN') are more frequent in terms of types than non-gradable adjectives (Class 1 'BLUE'). Therefore, this section addresses the question whether the *Semantic Complexity Hierarchy* is reflected in the adjective classes' mean age of acquisition

[10] A negative correlation between age of acquisition and token frequency can also be observed for the adjectives produced by Leo (r_s = -.765, $p < .001$). However, this is not surprising because for adjectives that are acquired early more occasions exist in which they can be produced.

(P1). The age of acquisition for each adjective was defined as the age of repeated use (see Stromswold, 1998). In addition, it was investigated whether the *Semantic Complexity Hierarchy* is reflected in the adjective classes' growth pattern (P2). Growth was understood as the age at which a semantic class "explodes", i.e., when the maximum number of new types is acquired.

4.3.2.1 Mean age of acquisition

The analysis regarding the acquisition order of semantic classes focuses on adjectives that are repeatedly used.[11] Overall, 154 of the 202 semantically analyzed adjectives were produced twice in one month or five times overall. As described in Section 4.3.1.3, non-subsective/non-intersective adjectives (Class 4 'FORMER') were not attested during the time period investigated. For all other classes, members of these classes were produced repeatedly already from 2;00. The mean and median age of acquisition including all members of the respective classes is given in Table 4.7.

Table 4.7: Leo: Age of repeated use in months.

Adjective class	*N*	Age range	Median age	Mean age (*SD*)
1 'BLUE'	28	24 - 35	28	28.18 (3.7)
2a 'CLEAN'	40	24 - 35	27	28.18 (3.7)
2b 'BIG'	34	24 - 33	25	26.41 (3.1)
3a 'CLEAN'	37	24 - 35	31	30.57 (3.0)
3b 'BEAUTIFUL'	15	24 - 35	28	28.60 (3.5)

Note. *N* = Number of types, *SD* = Standard deviation.

Table 4.7 shows two different measures of the frequency distribution's central tendency for the age of acquisition: the median and the arithmetic mean. The median is the middle age of acquisition when the ages of acquisition for each adjective are ranked in order of magnitude (Field, 2009), i.e., it divides the distribution into two halfs of the same size. The mean calculates the average age of acquisition. For the present purposes, the median is the more appropriate measure for two reasons. First, it is a more realistic measure of the age (in months): for instance, 28 months as the mean age of acquisition is more meaningful than 28.18 months. Second, the median is the "typical" age of acquisition of the distribution because it is less affected by extreme scores than the arithmetic mean (Field, 2009).

Regarding the question of whether the adjective classes differ in the age of acquisition (P1), two analyses were run. First, the age of acquisition for equally complex adjective classes was compared. This analysis tested the prediction that these classes should be acquired at the same age. If this prediction is borne out, equally complex adjectives can be analyzed as one class and in the second analysis the age of acquisition can be compared to

[11] The acquisition patterns are similar for the emergence of adjectives, i.e., the age of their first use. The results are given in Appendix A.5.

classes of different complexity. This way, it is tested whether adjective classes of different complexity are acquired at different ages.

The comparison of Class 2a ('CLEAN') and 2b ('BIG') and 3a ('HEALTHY') and 3b ('BEAUTIFUL') revealed, contrary to prediction, a significant difference in the age of acquisition between absolute and relative gradable dimensional adjectives (2a vs. 2b, Z = -2.268, p = .023) as well as between absolute and relative gradable evaluative adjectives (3a vs. 3b, Z = -2.011, p = .044). As a consequence, the two dimensional classes (2a and 2b) and the two evaluative classes (3a and 3b), respectively, cannot be collapsed. Collapsing these classes is necessary for the second analysis, i.e., comparing the mean age for non-gradable (1), dimensional (2), and evaluative (3) adjectives. Consequently, the age of acquisition for all five classes in Table 4.7 was compared. It was found that the age of acquisition is affected by semantic class ($\chi^2(4, \text{N} = 154) = 23.795$, $p < .001$). Post-hoc tests revealed that the mean age of acquisition differed significantly between absolute gradable evaluative (3a 'HEALTHY') and relative gradable dimensional (2b 'BIG') adjectives (z = 4.833, adjusted $p < .001$) as well as between absolute gradable evaluative (3a 'HEALTHY') and absolute gradable dimensional adjectives (2a 'CLEAN') (z = -2.843, adjusted p = .045). Due to the adjusted significance in multiple testing the difference between Class 2a and 2b and between 3a and 3b did not reach significance. As a result, more complex adjectives were acquired after less complex adjectives (2a $\stackrel{?}{=}$ 2b $<$ 3a) or concurrently with less complex adjectives (1 = 3a $\stackrel{?}{=}$ 3b, 2a $\stackrel{?}{=}$ 2b = 3b).

This order is based on adjective occurrences in all syntactic structures. How is the order influenced by the syntactic structures in which adjectives occurred? Several considerations suggest a more detailed analysis of the acquisition order by syntactic structure. First, at 2;00, members of all adjective classes, except for non-subsective/non-intersective adjectives, were attested. But as shown in Figure 4.1, Leo started to produce adjectives mostly in single-word utterances. Therefore, an acquisition order may be easier to detect in later acquired structures, i.e., when Leo starts to combine adjectives with nominal elements. Second, semantic properties become in particular relevant in compositional structures: entailment properties determine by which operation adjective and noun have to be combined. The noun also plays a role for the adjective's gradability because its denotation can provide the comparison class. In contrast, for notional classes, which are conceptually motivated, it is useful to analyze all occurrences because they target the composition of the lexicon instead of compositional semantics. Third, Figure 4.6 indicates a relation between semantic classes and the syntactic structure in which they are preferably produced. Therefore, it may be that the acquisition order differs depending on the syntactic structure. Hence, the mean age of acquisition is calculated separately for adjectives in attributive and predicative structures.

Eighty adjectives were produced repeatedly in attributive position. From 2;00, members of all adjective classes except for relative and absolute gradable evaluative adjectives (3a 'HEALTHY', 3b 'BEAUTIFUL') were productive in attributive position. Members of Class 3a and 3b were attested productively in attributive position from 2;01. The mean and median

age of acquisition including all members of the respective classes is given in Table 4.8.

Table 4.8: Leo: Age of repeated attributive use in months.

Adjective class	*N*	Age range	Median	Mean (*SD*)
1 'BLUE'	25	24 - 35	28	28.32 (3.5)
2a 'CLEAN'	12	24 - 35	30.5	30.33 (3.6)
2b 'BIG'	24	24 - 35	29	29.04 (3.8)
3a 'HEALTHY'	12	25 - 35	32.5	32.17 (2.8)
3b 'BEAUTIFUL'	7	25 - 35	30	29.57 (3.1)

Note. N = Number of types, SD = Standard deviation.

When used in attributive position, equally complex adjective did not differ significantly in their age of acquisition (2a 'CLEAN' vs. 2b 'BIG': Z = -.879, p = .380; 3a 'HEALTHY' vs. 3b 'BEAUTIFUL': Z = -1.883, p = .060). Therefore, they were analyzed together to assess whether more complex adjectives were acquired later than less complex adjectives or at the same age. The median age of acquisition for non-gradable (1 'BLUE') adjectives is 28 months, the median age of acquisition for dimensional adjectives (2a and 2b) is 30 months, and the median age of acquisition for evaluative adjectives (3a and 3b) is 32 months. The age of acquisition is distributed differently over non-gradable, dimensional, and evaluative adjectives ($\chi^2(2, \mathrm{N} = 80) = 6.616$, p = .037). Pairwise post-hoc tests revealed a significant difference in the mean age of acquisition between non-gradable and evaluative adjectives (z = -2.572, adjusted p = .030). As a result, more complex adjectives were acquired after less complex adjectives ($1 < 3$) or concurrently with less complex adjectives ($1 = 2$, $2 = 3$).

In predicative position, 91 adjectives were produced repeatedly. From 2;01, members of the class of relative gradable dimensional (2b 'BIG') and the class of relative gradable evaluative (3b 'BEAUTIFUL') adjectives were repeatedly used. From 2;02, members of the class of non-gradable (1 'BLUE') and the class of absolute gradable gradable evaluative (3a 'HEALTHY') were repeatedly used. Members of the class of absolute gradable dimensional adjectives (2a 'CLEAN') were repeatedly used from 2;03. The mean and median age of acquisition including all members of the respective classes is given in Table 4.9.

Table 4.9: Leo: Age of repeated predicative use in months.

Adjective class	*N*	Age range	Median	Mean (*SD*)
1 'BLUE'	13	26 - 31	28	28.0 (1.6)
2a 'CLEAN'	24	27 - 35	30	29.92 (2.6)
2b 'BIG'	22	25 - 34	28.5	29.27 (2.7)
3a 'HEALTHY'	23	26 - 35	30	30.61 (2.4)
3b 'BEAUTIFUL'	9	25 - 35	29	29.22 (2.9)

Note. N = Number of types, SD = Standard deviation.

When used in predicative position, equally complex adjective did not differ significantly in their age of acquisition (2a 'CLEAN' vs. 2b 'BIG': Z = -.689, p = .491; 3a 'HEALTHY' vs. 3b 'BEAUTIFUL': Z = -1.425, p = .154). Therefore, they were analyzed together to assess whether more complex adjectives were acquired later than less complex adjectives or at the same age. The median age of acquisition for non-gradable adjectives (1 'BLUE') is 28 months, the median age of acquisition for dimensional adjectives (2a and 2b) is 29 months, and the median age of acquisition for evaluative adjectives (3a and 3b) is 30 months. The age of acquisition is distributed differently over non-gradable, dimensional and evaluative adjectives (χ^2(2, N = 91) = 7.505, p = .023). Pairwise post-hoc tests showed a significant difference in the mean age of acquisition between non-gradable and evaluative adjectives (z = -2.739, adjusted p = .018). As a result, more complex adjectives were acquired after less complex adjectives ($1 < 3$) or concurrently with less complex adjectives ($1 = 2$, $2 = 3$). This means that the same order of acquisition is found for predicative and attributive adjectives.

In summary, regarding adjectives co-occurring with a nominal element, equally complex adjectives did not differ in their mean age of acqusition. For adjectives of different complexity, more complex adjectives were not acquired before less complex adjectives.

The data in Table 4.8 and Table 4.9 may be interpreted differently. Based on the median age of acquisition for each class one could argue that it is also possible to classify adjectives as absolute gradable (2a 'CLEAN' and 3a 'HEALTHY') and relative gradable (2b 'BIG' and 3b 'BEAUTIFUL') instead of classifying them as dimensional (2a 'CLEAN' and 2b 'BIG') and evaluative (3a 'HEALTHY' and 3b 'BEAUTIFUL'). Indeed, the latter classes do not differ in the mean age of acquisition either (2a vs. 3a: Z = -1.367, p = .172 (attributive), Z = -1.030, p = .303 (predicative); 2b vs. 3b: Z = -.190, p = .849 (attributive), Z = .000, p = 1.000 (predicative)). This classification would have consequences for the *Semantic Complexity Hierarchy*. I discuss these consequences in Section 4.4.

4.3.2.2 Growth patterns

In lexical acquisition the phenomenon of a so-called 'vocabuary spurt' or 'vocabulary explosion' has been reported (e.g., E. V. Clark, 1993; Menyuk et al., 1995). This phenomenon describes that children acquire a large amount of words in a short period of time. It is conceivable that this phenomenon also appears for specific word classes, in this case adjectives. With regard to the acquisition of adjectives, the present analysis examined how many new types Leo acquired across age. Figure 4.7 displays how many types are newly acquired in attributive and predicative position across age.

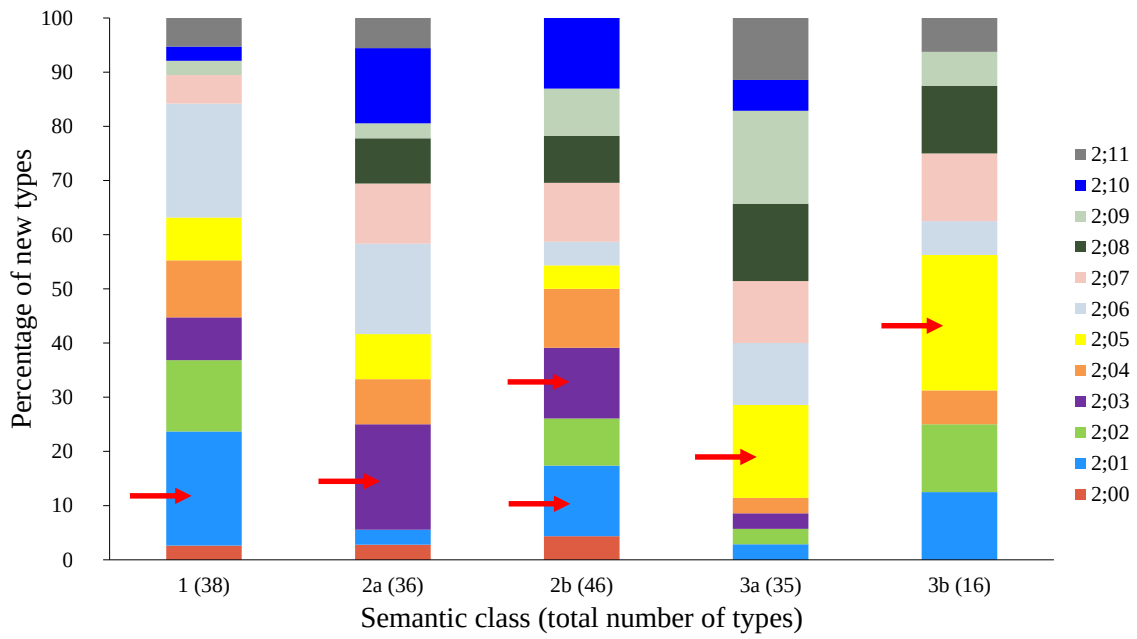


Figure 4.7: Leo: Percentage of newly acquired types in attributive position and predicative position across age. The arrows indicate the age at which the maximum number of new types is acquired.

It was analyzed at which age the number of new types reached its first maximum. As can be seen from the arrows in Figure 4.7, the age at which the number of new types reached its first maximum differs between adjective classes. The highest number of newly acquired types for non-gradable adjectives (1 'BLUE') was attested the first time at 2;01. At this age, eight adjectives were acquired. The highest number of newly acquired types for absolute (2a 'CLEAN') and relative (2b 'BIG') gradable dimensional adjectives was attested at 2;03 with seven new types for absolute gradable dimensional and six new types for relative gradable dimensionale adjectives. Note that for relative gradable dimensional adjectives (2b 'BIG') this was not the first time that six new types were acquired. Already at age 2;01, the same number of new types was acquired. The highest number of newly acquired types for absolute (3a 'HEALTHY') and relative (3b 'BEAUTIFUL') gradable evaluative adjectives was attested the first time at 2;05 with six new types for absolute gardable evaluative adjectives and 4 new types for relative gradable evaluative adjectives. In summary, adjective classes of the same complexity "exploded" at the same age: Class 2a and 2b at age 2;03, Class 3a and 3b at age 2;05. The class of non-gradable adjectives (1) "exploded" first.

4.3.2.3 Summary

In this section, I assessed hypothesis H1 that the order of adjective acquisition is determined by the adjectives' semantic complexity. The data indicate that the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition, measured as the median age of repeated use, and in the growth patterns for the semantic classes, measured as the highest number of newly acquired types: More complex adjective classes are not acquired before

less complex adjective classes and more complex adjective classes do not "explode" earlier than less complex adjective classes.

4.3.3 The role of the input

This section focuses on hypothesis H2 stating that the order of acquisition of adjectives is not completely determined by the input, in particular by the frequency of adjective use in the input. Previous findings regarding the role of the input for the order of adjective acquisition are mixed. On the one hand, the findings indicate a relation between adjective use in child speech and child-directed speech. On the other hand, it seems unlikely that input properties are the only predictor of the age of acquisition for three reasons. First, adjective use differed across caregivers, hence the children did not receive the same input, but the acquisition order for notional classes was similar across children and across different languages (Tribushinina et al., 2014). Second, caregivers used adjectives from notional classes that differed from the ones their children used (Blackwell, 2005). Third, it is open whether children's adjective use is a reaction on parental adjective use or *vice versa*. To test the hypothesis that the order of adjective acquisition is not completely determined by adjective use in the input this section investigates whether adjective frequency in child speech and child-directed speech are correlated (P3), whether adjective frequency in the input is reflected in the adjectives' age of acquisition for the adjective classes (P4), and whether adjectives that are acquired late are missing in the input before the age of acquisition (P5).

4.3.3.1 Adjective frequency in child speech and input

Leo's main caregivers, who provided the input to Leo, were his mother, his father, and a research assistant who helped with the recordings (see Behrens, 2006). Overall, Leo's caregivers produced 36,286 adjective tokens distributed over 882 adjective types. Out of the 882 adjective types that the CLAN command *freq*[12] revealed, 313 adjective types were attested both in Leo's speech and in the input.

Regarding (P3), it was analyzed whether the total number of tokens in child speech and in the input are correlated and whether the distribution of adjective tokens across semantic classes differed between child and parental use. These analyses are based on the 202 adjective types that were analyzed semantically (see Table 4.5). With respect to the total number of tokens, a positive correlation between child use and parental use is observed ($r_s = .737$, $p < .001$). The distribution of tokens across semantic classes is displayed in Table 4.10. The numbers for Leo's adjective use from Table 4.5 are repeated here.

[12] Version "freq (11-Apr-2017)"

Table 4.10: Distribution of tokens across semantic adjective classes in the input and in Leo's speech.

	Input		Leo	
Adjective class	*N* tokens	% tokens	*N* tokens	% tokens
1 'BLUE'	5264	18.4	3096	33.3
2a 'CLEAN'	2762	9.7	1312	14.1
2b 'BIG'	11,078	38.7	4017	43.2
3a 'HEALTHY'	1920	6.7	471	5.1
3b 'BEAUTIFUL'	7571	26.5	393	4.2
Total	28,595	100	9289	100

Note. *N* tokens = Raw number of tokens, % tokens = Percentage of tokens.

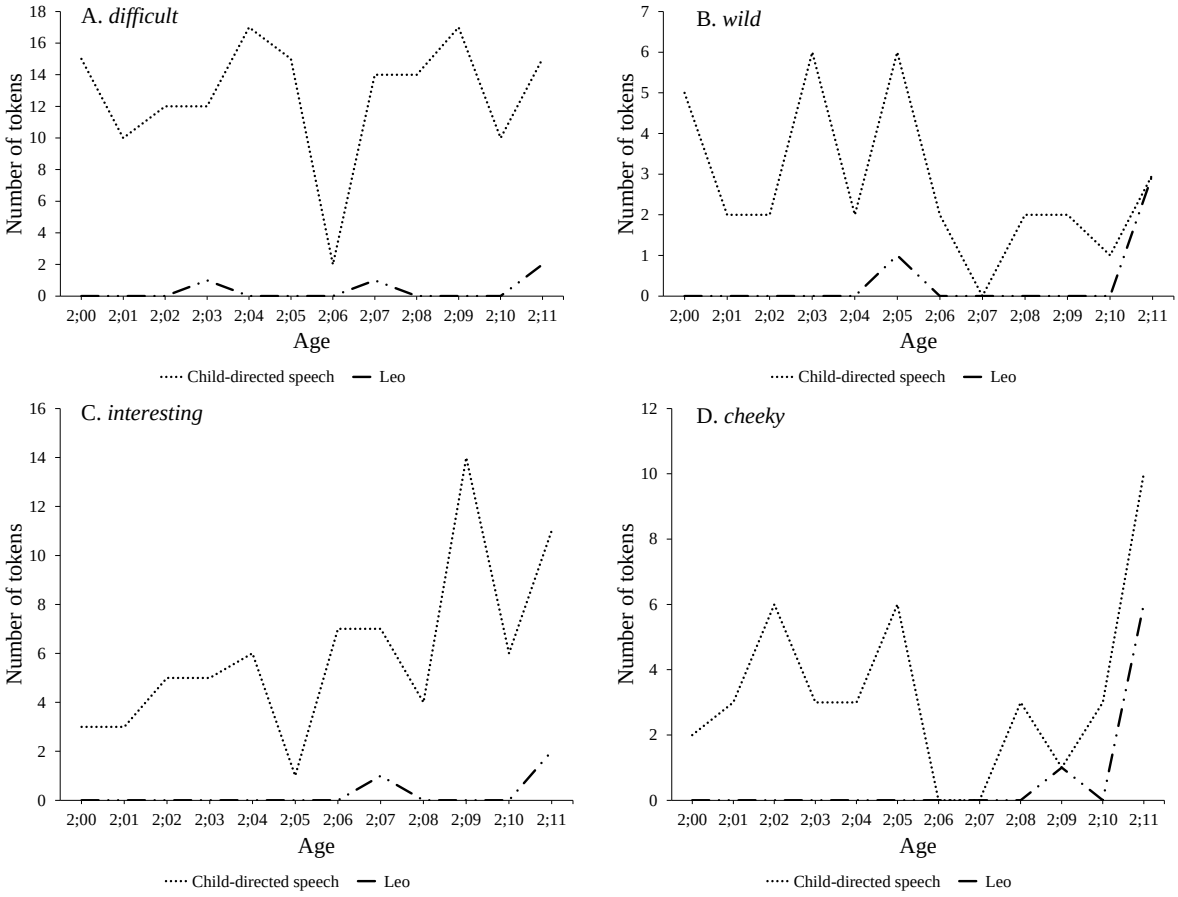
In the input the number of tokens is distributed differently across semantic classes ($\chi^2(4, N = 188) = 18.298$, $p = .001$; see Section 4.3.1.3 for the distribution in Leo's speech). Post-hoc tests revealed that the total number of tokens differed significantly between relative gradable dimensional (2b 'BIG') and absolute gradable evaluative (3a 'HEALTHY') adjectives (z = -3.058, adjusted $p = .022$) as well as between relative (2b 'BIG') and absolute (2a 'CLEAN') gradable dimensional adjectives (z = -3.441, adjusted $p = .006$). Recall that in Leo's speech relative gradable dimensional adjectives (2b 'BIG') differed from absolute and relative gradable evaluative adjectives (3a 'HEALTHY' and 3b 'BEAUTIFUL'). The findings indicate that relative gradable dimensional adjectives (2b 'BIG') adjectives are proportionally the most frequent class in both child speech and child-directed speech (see Table 4.10). Leo and his caregivers differ most clearly with regard to the token frequency of relative gradable evaluative adjectives (3b 'BEAUTIFUL'): in the input this is the second most frequent class, whereas in Leo's speech it is the least frequent class. In addition, differences can be observed for non-gradable (1 'BLUE') and absolute gradable dimensional (2a 'CLEAN') adjectives. Proportionally, these classes were used more frequently by Leo than by his caregivers.

4.3.3.2 Adjective frequency in the input and age of acquisition

Addressing (P4), it was analyzed whether the age of acquisition for the 202 adjective types that were analyzed semantically correlates with the total number of tokens for these adjectives in the input. Age of acquisition and total number of tokens in the input are negatively correlated (r_s = -.617, $p < .001$). Hence, one may conclude that the more frequent an adjective occurs in the input the earlier it is acquired. In other words, the negative correlation between age of acquisition and input frequency implies that the frequency of adjective use in the input is reflected in the adjectives' age of acquisition. As a consequence, late acquired adjectives should occur infrequently in the input. Are adjectives that children acquire late missing in parental speech before the age of acquisition (P5)?

To examine this question, the production of five adjectives in the input was analyzed:

schwierig ('difficult'), *interessant* ('interesting'), *wild* ('wild'), *frech* ('cheeky') and *hübsch* ('pretty'). These adjectives were selected because they were acquired by Leo at 2;11. For each of these adjectives, the utterances in which the adjectives occurred in the input were extracted via the CLAN command *kwal*[13]. Adverbial uses of these adjectives (see examples (4.4) and (4.5)) were excluded from the analysis. Note that for the input analyses, it was not differentiated between predicative and attributive adjective uses. Figure 4.8 compares the raw number of tokens in Leo's speech and in the input for the five selected adjectives.



[13] Version "kwal (11-Apr-2017)"

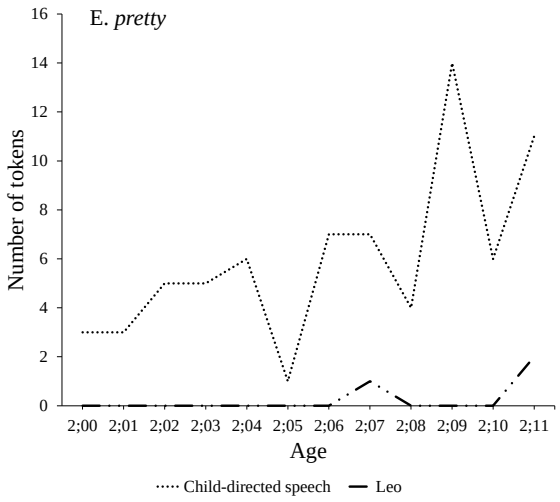


Figure 4.8: Raw numbers of tokens in the input and in Leo's speech.

As can be seen in Figure 4.8, these five adjectives were produced in the input from age 2;00. For three adjectives (*interessant, frech, hübsch*, see Figure 4.8.C/D/E) an increase in the number of tokens could be observed towards the end of the investigated time period, but for two adjectives (*schwierig, wild*, see Figure 4.8.A/B) this trend was not found. One could argue that although these five adjectives are attested in the input from 2;00, the number of occurrences is too low to enable an early acquisition. However, there are adjectives that are as infrequent as the above adjectives but were acquired early. *Leer* ('empty'), *offen* ('open'), *sauber* ('clean'), *trocken* ('dry'), *hart* ('hard') are among the adjectives that were acquired at 2;00. Table 4.11 illustrates that the total number of tokens is similar across these ten early and late acquired adjectives. More importantly, the number of tokens in the input at 2;00 is also similar.

Table 4.11: Token frequency of early and late acquired adjectives in the input.

Adjective	Total number of tokens	Number of tokens at 2;00
	Acquired at 2;11	
schwierig	153	15
wild	33	5
interessant	87	8
frech	40	2
hübsch	72	3
Total	385	33
	Acquired at 2;00	
leer	75	10
offen	19	6
sauber	42	11
trocken	61	8
hart	36	4
Total	233	39

In summary, adjectives that were acquired late were nevertheless present in the input across age. In addition, it has been shown that low frequency in the input does not prevent an adjective from being acquired early.

4.3.3.3 Summary

This section assessed hypothesis H2 stating that the acquisition order of adjectives is not completely determined by the adjectives' frequency in the input. The findings suggest (a) that adjective frequency in Leo's speech and in the input are positively correlated and (b) that adjective frequency in the input and age of acquisition of these adjectives are negatively correlated. However, it is open how the variables influence each other. With respect to (a), it is open whether the frequency of Leo's adjective use mirrors the frequency of his cargeivers' adjective use or *vice versa*. Regarding (b), it is possible that the child acquired an adjective early because it occurred frequently in the input. However, it is also conceivable that an adjective occurred frequently in the input because it was acquired early. Furthermore, I showed that adjectives with similar frequency in the input were acquired at different ages. In addition, the findings point to differences in adjective use between the child and his caregivers: some adjective classes were produced more frequently by the child than by his caregivers. Other classes were produced more frequently by the caregivers than by the child.

4.4 Development of the productive adjective lexicon

In this chapter I conducted three main analyses. In Section 4.3.1, the child corpus was described in terms of notional adjective classes, the development of the syntactic structures in which adjectives are used, and the relation between notional and semantic adjectives classes. Section 4.3.2 addressed the questions of whether the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition (P1) and in the growth patterns of the semantic adjective classes (P2). Section 4.3.3 addressed the questions of whether the frequency of adjective use in child and parental speech are correlated (P3), whether frequency of adjective use is reflected in the adjectives' age of acquisition (P4), and whether adjectives that children acquire late are missing in parental speech before the age of acquisition (P5).

Regarding the composition of the lexicon in terms of notional adjective classes, the findings for the present corpus of one German-speaking child between 2;00 and 2;11 years are consistent with previous findings for Dutch, English, French, Hebrew, and Turkish (Blackwell, 2005; Tribushinina et al., 2014). The present study extends previous findings on the development of adjectives in spontaneous speech by taking compositional and lexical semantic properties, i.e., entailment and gradability properties, into account. I showed that notional and semantic adjective classes are related although a one-to-one correspondence exists only for some classes: adjectives from the same notional class can have different semantic properties and adjectives from different notional classes can have the same semantic properties. Thus, the present study covers two aspects of lexical acquisition. First, the development of notional classes, which relate to the concepts adjectives express. Second, the development of semantic classes, which are particularly important when adjectives combine with nouns. Semantic classes provide information about the compositional and lexical semantic properties underlying the concepts expressed by notional classes.

With respect to the development of adjective structures, the data show that the child started to use adjectives in single-word utterances. Attributive and predicative structures emerged around the same age, at 2;01 years. However, attributive adjectives were produced more frequently than predicative adjectives. This distribution is consistent with the distribution of attributive and predicative adjectives found for English child language (Blackwell, 2001). These findings can serve as a starting point for further studies that may want to investigate whether differences in complexity also exist at a structural level.

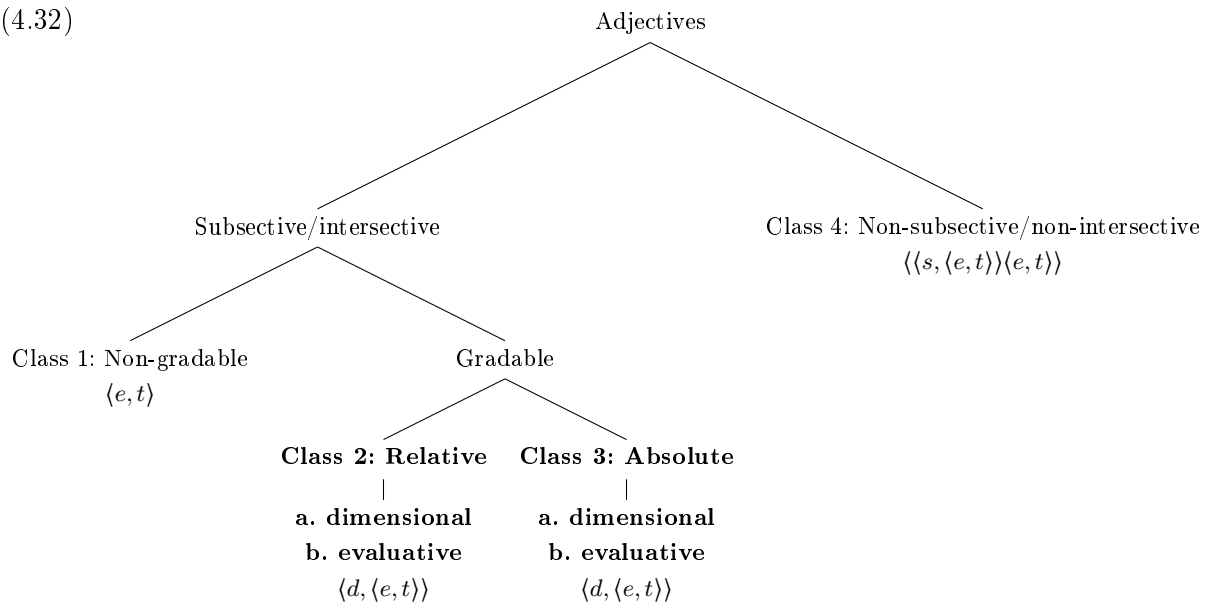
The data also point to a relation between syntactic position and both semantic and notional adjective classes: some adjective classes occurred primarily in attributive position whereas other adjective classes were more often attested in predicative position. Blackwell (2001) explains the interaction between syntactic and notional classes with the notion of "time stability" of the properties denoted by the adjectives (see Chapter 3). According to her account, more time-stable properties such as COLOR, DIMENSION, or VALUE are preferred in attributive position. In contrast, less time-stable properties such as MENTAL STATE, PHYSICAL STATE, or BEHAVIOR are preferably used in predicative position. The notion of

less and more time-stable properties resembles the stage-level/individual-level distinction. According to Toledo and Sassoon (2011), absolute gradable adjectives denote stage-level properties, whereas relative gradable adjectives denote individual-level properties. Hence, absolute gradable adjectives denote less time-stable properties and should be preferred in predicative position. Relative gradable adjectives denote rather time-stable properties and should be preferred in attributive position. The data of the present study confirm this prediction: Leo produced absolute gradable dimensional and evaluative adjectives more frequently in predicative position than in attributive position. Relative gradable dimensional adjectives were used more frequently in attributive than in predicative position. For relative gradable evaluative adjectives the distribution of attributive and predicative occurrences was balanced. For non-gradable adjectives it is difficult to categorize them as individual- or stage-level predicates. They occurred more frequently in attributive position. Hence, a more careful analysis is necessary to identify the reason for the interaction between the preferred syntactic position and the semantics of adjectives.

The data presented in Section 4.3.1.3 (see Table 4.5) suggest that neither type nor token frequency mirror semantic complexity. In Section 4.3.2, it was investigated whether the acquisition order of adjectives mirrors semantic complexity. Regarding (P1), the findings indicate that the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition for the adjective classes because adjective classes of the same complexity did not differ in their mean age of acquisition, whereas more complex adjectives were not acquired before less complex adjectives. This order of acquisition appeared most clearly in predicative and attributive structures. This finding mirrors that the compositional and lexical properties of the adjective classes concern the adjectives' function as modifiers of nominal elements. In other words, these properties are particularly important when adjectives occur together with nominal elements. The most complex class of non-subsective/non-intersective adjectives (e.g., *former*, *alledged*) was not attested until age 2;11. Regarding (P2), the findings suggest that the *Semantic Complexity Hierarchy* is also reflected in the growth patterns for the adjective classes. Adjective classes of the same complexity "exploded" at the same age. This means, at this age the highest number of new types was acquired. For adjective classes of different complexity, more complex adjective classes "exploded" later than less complex adjective classes. The outcomes of (P1) and (P2) are predicted by hypothesis H1 stating states that the acquisition order of adjectives is determined by their semantic complexity.

In Section 4.3.2, I remarked that the median age of acquisition of the semantic adjective classes may allow an alternative classification of adjectives, namely that absolute gradable adjectives may form one class, which is divided into dimensional and evaluative adjectives, and that relative gradable adjectives may form one class, which is also divided into dimensional and evaluative adjectives. As a consequence, the hierarchy would look like the following:

(4.32)



In Chapter 2, I noted that this branching is in principle possible, but that it does not follow from the definition of semantic complexity proposed in Chapter 2. Given the alternative hierarchy another definition of semantic complexity would have to answer the question why, according to the present results, absolute gradable adjectives should be more complex than relative gradable adjectives and why dimensional and evaluative adjectives should be equally complex. This alternative classification is suggested only by the age of acquisition but not by the growth patterns. What is more, the age of acquisition is also consistent with the originally proposed classification. Therefore, I conclude that the originally proposed classification in (4.1) should be maintained.

In general, the results for notional and semantic classes coincide. This finding points to a relation between these two classifications. In Section 4.3.1.3, I suggested how notional classes can be translated into semantic classes. The suggested relation between notional and semantic classes became apparent with respect to the distribution of types and tokens and with respect to the relation between syntactic position and adjective class. With regard to the acquisition order, notional and semantic classes are difficult to compare. The difficulty in comparing the acquisition order of notional and semantic classes results from the diversity of many notional classes: one notional class can include adjectives from different semantic classes. However, although it is difficult to identify a strict correpondence between the acquisition order for notional and semantic adjective classes, the following order can be observed: notional classes that correspond to gradable dimensional adjectives, i.e., DIMENSION, PHYSICAL PROPERTY, AGE, were acquired earlier than notional classes that correpond to gradable evaluative adjectives, i.e., VALUE, CONFORMITY, PHYSICAL STATE, MENTAL STATE, BEHAVIOR. This order is in line with the order predicted by semantic complexity. The order for notional classes that correspond to non-gradable adjectives is less consistent with the order predicted by semantic complexity because COLOR was the first notional class acquired, whereas NATIONALITY-adjectives were acquired rather late.

Section 4.3.3 focused on the role of the input for the acquisition of adjectives. Regarding (P3), the data showed a positive correlation between adjective frequency in child speech and in the input. This result suggests a general influence of the adjective use in the input on the child's adjective use. However, type and token frequency were distributed differently across adjective classes and child and parental speech indicating an independent adjective use in child and parental speech. Regarding (P4), adjective frequency in child-directed speech was negatively correlated with the age of acquisition suggesting that the acquisition order is influenced by the frequency of adjective use in the input. Regarding (P5), I showed by means of specific adjectives that a late or an early age of acquisition is independent of the frequency of use for an adjective in the input. The outcomes of (P3), (P4), and (P5) are predicted by hypothesis H1 because they indicate that adjective frequency in the input does not completely determine the acquisition order of adjectives. Nevertheless, to finally clarify the role of the input for the acquisition of adjectives a more detailed analysis is necessary. Further research on the input should be extended to a broader range of adjectives. Moreover, so far only token frequency for adjectives that were produced by the child and his caregivers were analyzed. But as shown in Section 4.3.3.1, more types were found in the input than in the speech of the child. To examine whether adjectives with the same semantics are used in child speech and in the input, the caregivers' adjectives must be classified according to the criteria in (4.31), too. What is more, as noted by Tribushinina et al. (2014), it cannot be concluded from correlation analyses that adjective use in the input influences adjective use in child speech because one "cannot be confident that it is parents rather than children who take the lead in increasing adjective use" (Tribushinina et al., 2014, p. 214). Therefore, a careful analysis of whether an utterance by the caregivers precedes or follows an utterance by the child containing the same adjective would be beneficial to this question. In addition, the structures in which adjectives occur should be examined in more detail: Is the distribution of attributive and predicative adjective uses similar in child speech and child-directed speech? Does the child receive positive evidence that adjectives denote properties of individuals? The latter question relates to the role the noun plays in the acquisition of adjectives. As reported in Chapter 3, Mintz and Gleitman (2002) found that two-year-old children understand novel words more easily as adjectives if the novel word modifies a taxonomically specific noun (e.g., 'Look at this *stoof* horsie. This horsie is very *stoof.*') compared to for instance pronouns (e.g., 'Look at this *stoof* thing. This one is really *stoof.*'). Findings by Blackwell and Olson (2008) and Sandhofer and Smith (2007) indicate that caregivers use adjectives also in other contexts. If this would also be the case in German, it is questionable whether the child can use the input to infer the meaning of adjectives. An additional challenge for the acquisition of German adjectives is their function as adverbs while maintaining the same form (as in predicative structures). In example (4.33a) *schön* ('beautiful') describes an individual, in example (4.33b) *schön* ('beautiful') describes an action.

(4.33) a. Schneewittchen ist schön.

'Snow White is beautiful.'

b. Schneewittchen singt schön.
'Snow White sings beautifully.'

Because the same form does not always denote a property of individuals, a one-to-one-correspondence between form and function does not exist, which makes it even more difficult to infer the meaning of adjectives from their use in the input.

In summary, the longitudinal analysis of the adjective use of one German-speaking child between 2;00 and 2;11 years showed that the acquisition order of adjectives is influenced by the adjectives' semantic complexity. This finding extends previous research on the acquisition of notional classes, i.e., the conceptual development of the word class ADJECTIVE, by taking compositional and lexical semantic properties of adjectives into account. These properties are crucial for the occurrence of adjectives in combination with nouns. The present preliminary analyses of adjective use in the input revealed that lexical acquisition is not entirely independent from the input, but that the input cannot be the leading factor in the acquisition of the semantics of adjectives. To generalize the results, data from more children and a more detailed analysis of the input are neccesary.

Spontaneous speech data allow only indirect inferences about the comprehension of adjectives. By investigating the occurrence of adjectives in comparison constructions or in combination with degree modifiers, previous studies conclude that the acquisition of (gradable) adjectives is a long process although they are produced early in spontaneous speech (Hohaus et al., 2014; Tribushinina & Gillis, 2012). The next chapter investigates the comprehension of two kinds of adjectives: relative and absolute gradable dimensional adjectives. These adjectives are assumed to be equally complex and the present study implied that they do not differ in their mean age of acquistion in production. Children's interpretation of these two adjective classes is important for the notion of semantic complexity because it is open whether children know the differences in meaning between the two classes regardless of the classes' equal complexity.

Chapter 5

Study 2: The comprehension of relative and absolute gradable adjectives

In Chapter 2, I stated in hypothesis H1 that the order of acquisition of adjectives is determined by their semantic complexity. The findings of the spontaneous speech data analysis reported in Chapter 4 provide first evidence for this hypothesis regarding the production of adjectives. However, spontaneous speech data allow only limited inferences regarding the interpretation of the adjectives attested in children's productive vocabularies. The experiment reported in this chapter investigated children's comprehension of adjectives, more precisely absolute gradable dimensional adjectives (AG) – Class 2a ('CLEAN') in the *Semantic Complexity Hierarchy* – and relative gradable dimensional adjectives (RG) – Class 2b ('BIG') in the *Semantic Complexity Hierarchy*.[1] The relevance of RGs and AGs for the notion of semantic complexity in acquisition is twofold. First, according to my definition of semantic complexity proposed in Chapter 2 they are supposed to be more complex than non-gradable adjectives – Class 1 ('BLUE') in the *Semantic Complexity Hierarchy*. As outlined in Section 2.2.3, I follow an analysis of gradable adjectives that assumes that they denote relations between individuals and degrees (type $\langle d, \langle e, t \rangle \rangle$). Consequently, they are more complex than non-gradable adjectives that denote properties of individuals (type $\langle e, t \rangle$). Second, I claim that RGs and AGs are equally complex. However, gradable adjectives show differences in the interpretation of the positive unmarked form regarding the standard of comparison and the relevance of the comparison class (see Section 2.2.3) although they share the same semantics. The interpretative differences can be attributed to lexical semantic properties such as the adjectives' scale structure (see Kennedy, 2007; Kennedy & McNally, 2005). The present study investigated whether children interpret all gradable adjectives alike due to their identical semantic type or whether they include the adjectives' lexical semantic properties, and hence distinguish between RGs and AGs.

[1] Parts of the experiment have been published in Weicker and Schulz (2018).

As pointed out in Chapter 3, previous research on the acquisition of gradable adjectives focused mainly on RGs; studies comparing the comprehension of RGs and AGs are scarce. The experiment presented in this chapter extends previous research on gradable adjectives by examining positive and negative pole RGs (*groß* 'big', *klein* 'small') and AGs (*sauber* 'clean', *dreckig* 'dirty'), and by investigating the influence of explicit comparison classes on the interpretation of RGs and AGs. Investigating RGs and AGs of different polarity can contribute to the theoretical question whether the standard of comparison must be taken to be a single point on the respective scale or a range of degrees (see Section 2.2.3). In contrast to previous studies (Syrett, 2007; Syrett et al., 2006, 2010), the participants of the present study were presented with both positive and negative pole adjectives. This way, it could be examined whether children divide the scale completely between the antonym pairs or whether they exhibit a 'gap' between the negative and the positive antonym.

In addition, the present study extends research by Syrett (2007), Syrett et al. (2006), and Syrett et al. (2010) by examining how children interpret RGs and AGs when the comparison class is linguistically encoded by the noun, i.e., when it is explicit. In other words, can children use the information provided by the noun to determine the comparison class? And for which kinds of gradable adjectives does the comparison class influence the standard of comparison? To answer these questions a design similar to Barner and Snedeker's (2008) study was used. They manipulated the comparison class by using different nouns, i.e., different explicit comparison classes, without changing the visual context. However, the original design was modified in five crucial aspects. First, the present study investigated both positive and negative RGs and AGs rather than RGs only. Second, a within-subject design was used instead of a between-subject design. The advantage of a within-subject design is that it can assess whether differences in the comparison class, i.e., the noun, influence the participant's interpretation of the adjective for the same visual display. In contrast, Barner and Snedeker (2008) compared the standards of two groups who saw the same array of objects, but heard different nouns. Third, in the present experiment the adjectives modified existing nouns instead of novel nouns. The third modification is related to the fourth modification that the age range of the participants was extended to younger and older children. Barner and Snedeker (2008) tested 4-year-old children, in the present experiment 3- to 5-year-old children were tested. Using existing nouns is relevant to examine whether knowledge about existing objects affects the interpretation of gradable adjectives. Barner and Snedeker (2008, p. 606) mention that "[f]uture studies should examine whether there is a stage at which very young children have knowledge about typical sizes of things in absence of a capacity to deploy a standard of comparison for novel sets. Here, we have shown that 4-year-olds have combined these two forms of knowledge [...]", but they did not test children younger than age 4. Using known instead of novel objects is a first step towards a better understanding of children's developmental stages regarding the comprehension of gradable adjectives. The fifth modification concerns the way the objects in the visual context were presented. In the present study, the distribution of objects differed across trials to test whether the participants can adjust the standard accordingly. In the

relevant experiments described in Barner and Snedeker (2008, Experiments 3 and 4) the distribution of objects was the same for all trials. What is more, in the present study the objects were presented without any ordering with the consequence that the participants had to establish their own ordering before calculating the standard of comparison. This way, the participants were prevented to infer their judgements from an already given order (cf. Foppolo & Panzeri, 2013; Syrett, 2007; Syrett et al., 2006, 2010). In addition, in previous studies that investigated whether children are able to shift the standard of comparison (Ebeling & Gelman, 1988, 1994; Syrett, 2007; Syrett et al., 2006, 2010) children often had to compare only two objects from the same basic-level category. In the present experiment, children had to compare eight objects from one basic-level category (Part 1) and twelve objects from two different basic-level categories (Part 2 and 3) (see Barner & Snedeker, 2008).

This chapter is organized as follows: in Section 5.1, I repeat the research questions and expected outcomes under hypothesis H1. Sections 5.2 and 5.3 describe the sample of participants and the practice trials of the experiment. The method and the results of the first part of the experiment, which address the first research question, are described and discussed in Section 5.4. The method and the results of the second and third part, which address the second research question, are described and discussed in Section 5.5. The chapter closes with a description of the developmental steps in the comprehension RGs and AGs suggested by the experimental findings.

5.1 Research questions

The present comprehension study examined whether and at which age pre-school children become sensitive to the distinction between relative and absolute gradable adjectives with respect to the standard degree and the relevance of the comparison class. The following research questions are addressed.

(C1a.) Do monolingual German-speaking children differentiate between absolute and relative gradable dimensional adjectives when defining the standard of comparison?

(C1b.) Does the standard of comparison change with age?

(C2a.) Which gradable dimensional adjectives do monolingual German-speaking children interpret relative to a comparison class?

(C2b.) Do interpretation patterns with respect to the relevance of the comparison class change with age?

Hypothesis H1 introduced in Chapter 2 states:

(H1) The order of acquisition of adjectives is determined by their semantic complexity.

Hypothesis H1 predicts that RGs and AGs are acquired, i.e., interpreted adult-like, at the same age due to their identical complexity. If children differentiate target-like between RGs and AGs, the following responses are predicted (e.g., Kennedy & McNally, 2005): regarding research question (C1a.), the standard for RGs should be located around the center of the scale. For AGs, the standard should be either a non-zero degree of the scale for minimum standard gradable adjectives such as *dreckig* ('dirty') or the maximal degree of the scale for maximum gradable adjectives such as *sauber* ('clean'). A related theoretical question is how the standard can be defined. According to Solt (2011), the standard for RGs should be analyzed as a range of degrees rather than a single point on the respective scale (see Section 2.2.3). The difference between the two analyses becomes apparent in the case of antonym pairs such as *big* and *small*. If the negative antonym is analyzed as the direct oppsosite of the positive antonym, the standard for both is a single point that completely divides the scale into big and small entities. But as Solt (2011) points out there are sizes that are judged as neither big nor as small (so-called 'borderline-cases'). As a consequence, one would have to assume two different standards for *big* and *small* with a gap in between. Crucially, the standard for *big* should be higher than the one for *small*, but according to Solt (2011) there should be nothing that rules out the possibility that the position of the two standards can be reversed. This means the standard of the negative antonym would be at a higher position on the scale than the standard of the positive antonym. As a result, sentences like *Dumbo is both tall and short for an elephant* should be true in some contexts. Defining the standard as range rather than a single point can account for these observations. Unlike in RGs, borderline-cases do not exist for AGs. Therefore, AGs are expected to exhibit a standard that is a single degree shared by the positive and the negative antonym. However, Rotstein and Winter (2004) illustrate that absolute gradable antonyms are not always interpreted as direct opposites of each other. In example (2.58), repeated here as (5.1), the degree of dirtiness of the glass is judged as neither clean nor dirty.

(5.1) This glass is almost dirty: It is certainly not clean, since it has some small spots on it, but it is not really dirty, and I am willing to drink from it if you insist.

Nevertheless, according to Rotstein and Winter (2004), the default meaning is to interpret absolute gradable antonym pairs as direct opposites. Hence, it is expected that participants should judge antonym pairs as direct opposites for AGs, but not for RGs.

Regarding research question (C2a.), participants should interpret RGs, but not AGs, relative to the comparison class if they differentiate target-like between RGs and AGs. Because AGs have (partially) closed scales, the endpoints can serve as the standard of comparison. Thus, AGs are context-independent and the value of the comparison class is not relevant for calculating the standard. In contrast, RGs have open scales without endpoints. Therefore, the standard for RGs is context-dependent and the value of the comparison class is relevant for calculating the standard (see Kennedy & McNally, 2005). In the present experiment the comparison class was explicit, i.e., expressed by the noun,

and not implicit, i.e., provided by the context (von Stechow, 1984). Hence, participants should adjust the standard according to changes in the noun for RGs, but not for AGs.

This experiment also investigated whether there is a developmental stage in which children do not have an adult-like interpretation of gradable adjectives (C1b. and C2b.) and how the non-adult-like interpretation can be characterized exactly. Due to the identical complexity of RGs and AGs it is likely that initially, children have the same interpretion for RGs and AGs. As intrdoduced in Chapter 2, the same interpretation of RGs and AGs can show up in two different response patterns. First, either relative or absolute gradable adjectives may be selected as the default interpretation. Second, it is possible that children do not have a default interpretation as findings by Syrett (2007) suggest, and thus neither relative nor absolute gradable adjectives may be interpreted adult-like initially. For this pattern, it is open how children interpret gradable adjectives instead. The first pattern – the default interpretation – means that either both RGs and AGs are interpreted relative to the comparison class and with a standard around the center of the scale or both RGs and AGs are interpreted independent of the comparison class and have a standard at one of the endpoints of the scale. For this pattern, it is open whether absolute or relative gradable adjectives are more likely to be the default interpretation. Tribushinina (2013) showed for RGs that 3-year-old children attach *big* and *small* only to the extremes of the range of objects, i.e., to the smallest or biggest object. Gao et al. (2014) found that 3- and 4-year-old children do not interpret adjectives relative to the noun if the visual context is the same. These findings point to an AG default interpretation. In contrast, as shown by Syrett et al. (2006) and Syrett et al. (2010) 3-year-old-children do not always interpret the AG *full* as having a maximum standard. This finding points to an RG default interpretation.

To test the predictions and answer the questions in (C1) and (C2), children of different ages were tested. The sample is described in the next section. Each experimental session was divided into three parts. Part 1 addresses (C1), Part 2 and Part 3 address (C2). Part 1 is described and discussed in Section 5.4, Part 2 and 3 in Section 5.5. In Section 5.6, I relate the findings of the different parts.

5.2 Participants

Forty-three monolingual German-speaking children from three age groups participated in the experiment: 11 three-year-olds (5 girls, 6 boys, age range: 3;2 to 3;11 years, mean age: 3;7 years), 15 four-year-olds (7 girls, 8 boys, age range: 4;1 to 4;11 years, mean age: 4;6 years) and 17 five-year-olds (9 girls, 8 boys, age range: 5;0 to 5;9 years, mean age: 5;4 years). Nine additional children were excluded because they were either in fact bilingual, did not complete the task or had a T-value below 40 in two or more subtests of the SETK 3-5 (Grimm, 2001), a standardized test that was used to ensure the children's typical language development. All children were tested at their day-care centers in the Frankfurt area. Twenty-six adult native speaker of German (22 female, 4 male) served as controls. The adults were undergraduate students of Goethe-University Frankfurt with little or no

linguistic background. They received either 10€ or credit points for participation. Because a within-subjects design was used, every participant completed every trial described in Section 5.4 and 5.5.

5.3 Practice trials

Three practice items without adjectives, e.g., *Gib mir bitte die Teddies* ('Please give me the teddys'), served to familiarize the participants with the task. The items are listed in Appendix B.1. In addition, it was checked whether participants knew the objects and their names that were used in the experiment. Participants received corrective feedback for the vocabulary check and the practice trials.

5.4 Part 1: Standard of comparison

In this section, the first part of the experiment is described. This part addresses the question whether and at which age children calculate different standards of comparison for AGs and RGs (C1a.) and whether the standard of comparison changes with age (C1b.). It is also analyzed whether the standard for RGs and AGs is defined as a range or a point on the scale. Based on previous studies (Syrett, 2007; Syrett et al., 2006; Tribushinina, 2013), it is expected that children differentiate between RGs and AGs, but that their standard of comparison shows some development before they have an adult-like interpretation.

5.4.1 Method

5.4.1.1 Materials

A forced picture-choice task inspired by Barner and Snedeker's (2008) design was used to investigate how children calculate the standard of comparison for RGs and AGs. In each trial, eight depicted objects from the same basic-level category were presented simultaneously (see Figure 5.1).[2] In contrast to previous studies (Foppolo & Panzeri, 2013; Syrett, 2007; Syrett et al., 2006), the objects came in unordered fashion to induce the participants to establish their own ordering. This way of visual presentation should prevent the participants to infer their judgements from an already given order.

[2] Note that the numbers on the picture cards were not present in the original experiment. They are added in Figure 5.1 for easier reference to the picture cards in the result section.

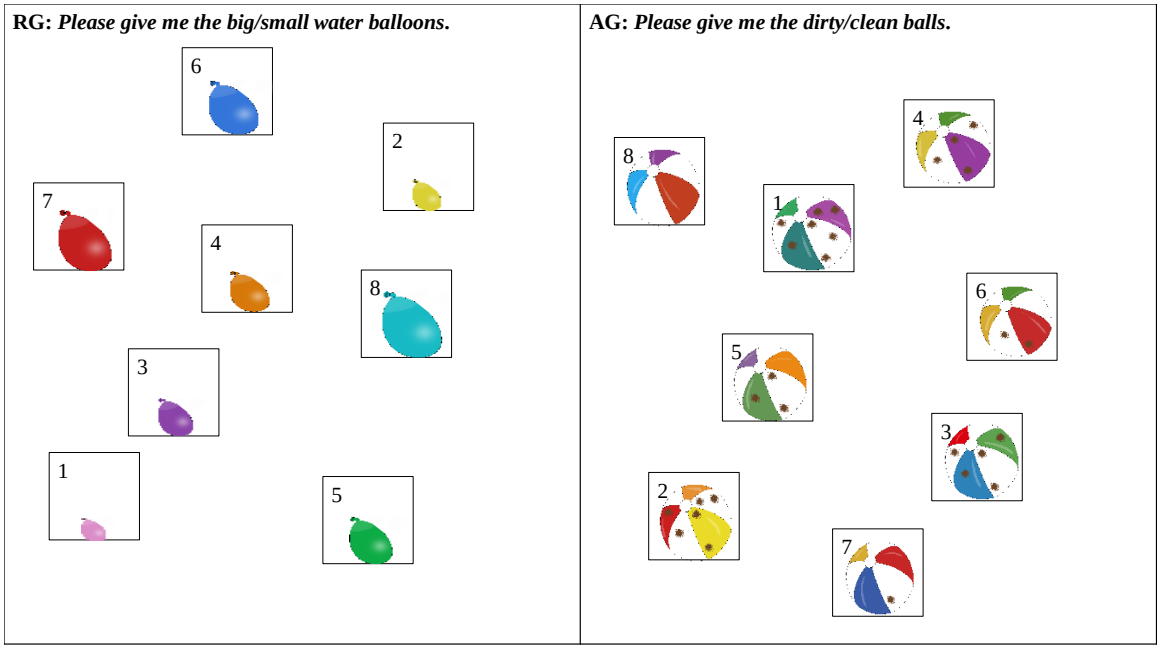


Figure 5.1: Example test trials for relative gradable adjectives (left) and absolute gradable adjectives (right).

Unlike in previous studies, participants were not asked to attest a respective property for every object individually. Instead, the participants' task was to select the objects matching their interpretation of the test prompt. The test prompts of the present experiment were always of the form *Gib mir bitte die Adj* N_{Plural} ('Please give me the Adj N_{Plural}'), e.g., *Gib mir bitte die großen Wasserbomben* ('Please give me the big water balloons'). The test prompts were uttered with a neutral, i.e., non-contrastive, intonation. This kind of test prompt was chosen for two reasons. First, in Part 2 and 3 (see Section 5.5) the noun played an important role. This way of asking allowed for a consistent format of the test prompts in all parts of the experiment. Second, pilot studies revealed that asking to attest a respective property for every object individually was tiring, in view of the fact that the trials in Part 2 and 3 of the experiment consisted of more than eight objects. The practice trials, which participants received before Part 1, illustrated that the participants were allowed to select as many or few objects as they wanted. That is, selecting only one object, although the noun in the test prompt is a plural DP, was introduced as an appropriate response. If the participants noticed that only one object was available matching the test prompt, the experimenter explained that the puppet's request sounded the same independent of the number of objects.

In total, Part 1 of the experiment comprised eight test items (= two items per adjective) and eight filler items. All test and filler items and the corresponding pictures are displayed in Appendix B.2. Two relative gradable adjectives (positive pole: *groß* 'big', negative pole: *klein* 'small') and two absolute gradable adjectives (positive pole: *sauber* 'clean', negative pole: *dreckig* 'dirty')[3] were tested. *Big* and *small* were used as RGs because previous studies

[3] In the remainder of this dissertation I use the English translations for the relevant adjectives and

(Bartlett, 1976; Eilers et al., 1974) showed that these less specific dimensional adjectives are acquired before more specific dimensional adjectives such as *long* or *narrow*. *Clean* and *dirty* were chosen as AGs because they were easy to illustrate and they were suitable to modify the objects which were used in the task. The materials consisted of picture cards (14x14 cm) that each displayed one object. All objects depicted belonged to the superordinate-level category "toys": water balloons, space hoppers, teddy bears and balls. Every object was of a different color. The property denoted by the adjective was realized to different degrees across the objects, i.e., they changed from small to big or from dirty to clean as illustrated in Figure 5.1.

The objects were chosen such that they could hold the adjectival property to varying degrees. Furthermore, the objects should be able to exhibit a natural transition from small to big or from dirty to clean. In addition, the aim was to avoid using entities for which the relative scale of these entities could interfere with the children's interpretation of gradable adjectives.[4] Toys can exhibit a natural transition from small to big or from dirty to clean, and they are entities that can be depicted as relatively close to the size of their referents in reality. Moreover, the toys used in this experiment are such that they are not miniatures of entities which are naturally large like for example toy cars. Another reason why toys were chosen becomes apparent in Section 5.5.

Importantly, the visual contexts were identical for *big* and *small*, and for *clean* and *dirty* respectively. Thus, it was possible to analyze whether the participants interpret negative and positive antonyms as direct opposites or whether there are some objects that are interpreted as exhibiting neither the property denoted by the positive nor by the negative antonym, e.g., objects that count as neither big nor small.

The filler trials consisted of eight picture cards displaying toys (buckets, dices, soccer balls, books, and Lego® bricks). In each trial, objects from two different basic-level categories were shown. The objects differed in shape (round or square) and color (blue or red) as exemplified in Figure 5.2. The filler items corresponded to the trials in which the participants had to make a request themselves. The requests were similar to the test prompts, but used different adjectives (shape or color), e.g., *Gib mir bitte die blauen Spielsachen* ('Please give me the blue toys').

nouns.

4 Huang and Snedeker (2013, p. 1093) point out that the depicted objects should be consistent with their real-world size and that miniature models of objects that are ordinarily quite large may interfere with children's interpretation of adjectives. Tribushinina (2013) reports that adults and 7-year-old children sometimes refused to call the test objects *big* because they were all smaller than in reality. This confirms Huang and Snedeker's (2013) concern.

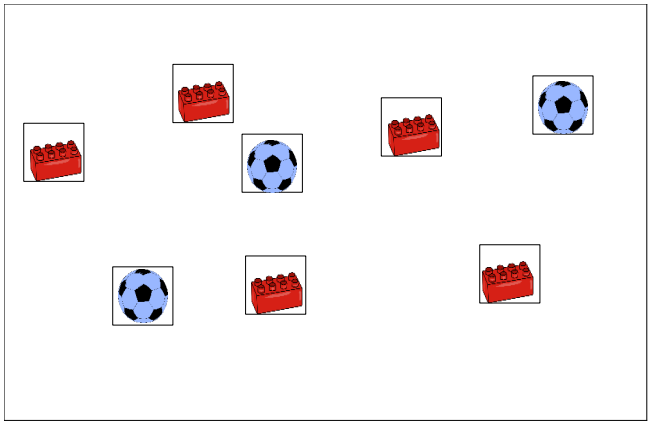


Figure 5.2: Example filler trial.

The filler items were particularly relevant for Part 2 and 3 of the experiment. Because a within-subjects design was used, participants saw the same visual context twice, but with different test prompts. The filler items served to enlarge the distance between the trials using the same visual context in order to reduce possible influences of the first presentation.

5.4.1.2 Procedure

The testing was divided into two sessions due to limited attention capacities of the younger children. Together with Part 2 and 3 and the standardized test each session lasted about 20 to 30 minutes. In the first session the positive adjectives *big* and *clean* were tested, in the second session their negative counterparts *small* and *dirty* were examined. The visual stimuli and the order of presentation were kept identical for both sessions to compare the relation between positive and negative adjectives directly. The sessions were about 12 days apart to minimize the influence of the items from the first session on the items in the second session. The children were tested individually in a quiet room at their day-care centers. The adults were tested at the university. The participants sat next to the experimenter on the floor or at a table that was large enough to show all picture cards simultaneously.

At the beginning of each test session the participants were introduced to a hand puppet called Adi. The participants were told that the puppet wants to play a game for which she brought some cards. The puppet and the participant each got one special dice. Both had to roll the dice without letting the other one see the outcome and then they had to make a request. The puppet's requests corresponded to the test trials, and the participant's requests corresponded to the filler trials. This procedure engaged the participants in a natural game situation. The participant's dice showed either a square, or a circle, or a blue color dot, or a red color dot. When it was the participant's turn, the experimenter distributed the picture cards, the participant rolled the dice and made a request. When the dice showed for instance the blue dot, the participants had to ask the puppet for the blue toys. The puppet selected the matching objects and handed them to the participant. When it was the puppet's turn, the experimenter distributed the picture cards, the puppet

rolled the dice and made a request equating the test items. The participants had to select the objects that in their opinion matched the request and handed them to the puppet. The experimenter emphasized that the winner of the game is the person who selected the correct picture cards more often. Otherwise, the children may have thought that the winner is the person with the most picture cards, perhaps influencing the children's choices. Test items and filler items were intermixed resulting in two test trials followed by two filler trials etc. Both sessions started with a trial containing a relative gradable adjective and then alternating absolute and relative gradable adjectives. At the end of Part 1 the participant was told that the puppet needs a break. During the puppet's break the participant was tested with an unrelated task before turning to Part 2 and 3.

5.4.1.3 Data analysis

To address the question whether children differentiate between RGs and AGs when defining the standard of comparison, it was analyzed which objects the participants considered big, small, clean, and dirty, respectively. At the group level, the data were analyzed in terms of the percentage of choices for each object across the two trials for the respective adjective. This analysis yielded the location of the standard of comparison on the scale for *big, small, clean* and *dirty*.

At the individual level, the standard of comparison was coded as the cut-off point. For *big* this was the smallest object considered big (together with all bigger objects), for *small* the biggest object considered small (together with all smaller objects), for *clean* the dirtiest object considered clean (together with all cleaner objects), and for *dirty* the cleanest object considered dirty (together with all dirtier objects). Consider for example the item on the left side in Figure 5.1: if a participant selected for instance object 6, 7 and 8, then the cut-off point for *big water balloons* was 6 marking the rank of the transition between big and non-big objects. The same objects were also presented asking for *small water balloons* in the second session. If a participant selected object 1, 2 and 3, then the cut-off point for *small water balloons* was 3 marking the rank of the transition between small and non-small objects. A response was coded as unanalyzable if participants selected for instance object 6 and 8 when asked for *big water balloons*, but did not select object 7. By comparing the cut-off points for the positive and the negative adjective it was possible to analyze whether the negative and the positive adjectives are direct opposites or whether there are some objects that refer neither to the positive nor the negative adjective.

The cut-off points of the child and adult group were statistically compared by Mann-Whitney-U tests. The cut-off points of the 3-, 4-, and 5-year-old group were statistically compared by Kruskal-Wallis tests. In the following, test statistics are only reported for significant results. All other measures are reported in Table B.2 and B.3 in the Appendix.

Recall that a target-like response for RGs was to locate the cut-off point around the center of the scale with a gap between the cut-off point for positive and negative antonyms. For AGs either the object which exhibits the adjectival property to the maximum (for

clean) or the minimum (for *dirty*) degree was the target-like cut-off point resulting in no gap between the cut-off point for positive and negative antonyms.

5.4.2 Results

5.4.2.1 Group data

This part of the experiment comprised two trials per adjective.[5] Table 5.1 summarizes the mean cut-off point (m), the central cut-off point (median = M) and the object that constitutes the cut-off point most often (modus = D) for each trial.

Table 5.1: Cut-off points for each test trial per age group.

Age		Cut-off point							
		big water balloons	*big space hoppers*	*small water balloons*	*small space hoppers*	*clean teddies*	*clean balls*	*dirty teddies*	*dirty balls*
3	m	4.2	4.9	3.1	3.2	7.8	7.2	7.0	6.5
	SD	2.0	1.5	1.1	1.0	0.4	2.0	0.0	1.6
	M	4.5	5.0	3.0	3.0	8.0	8.0	7.0	7.0
	D	1.0	5.0	2.0	2.0	8.0	8.0	7.0	7.0
4	m	5.6	5.5	3.1	2.8	7.5	7.1	7.1	7.1
	SD	1.3	1.6	1.3	1.3	1.9	2.5	0.3	0.3
	M	6.0	6.0	3.0	2.5	8.0	8.0	7.0	7.0
	D	6.0	7.0	2.0	2.0	8.0	8.0	7.0	7.0
5	m	5.5	5.3	3.0	3.4	7.9	7.9	7.0	7.0
	SD	1.3	1.2	1.1	1.0	0.3	0.3	0.0	0.4
	M	5.0	5.0	3.0	4.0	8.0	8.0	7.0	7.0
	D	5.0	5.0	2.0	4.0	8.0	8.0	7.0	7.0
Adults	m	5.9	6.2	3.0	3.2	8.0	8.0	7.0	7.0
	SD	0.7	0.8	0.9	0.9	0.0	0.0	0.0	0.0
	M	6.0	6.0	3.0	3.0	8.0	8.0	7.0	7.0
	D	6.0	6.0	3.0	3.0	8.0	8.0	7.0	7.0

Note. m = mean, SD = standard deviation, M = median, D = modus.

Seventeen of 552 answers were unanalyzable, hence the standard could not be determined. Therefore, they are missing in the calculations in Table 5.1. The graphs in Figure 5.3 to 5.6 take all answers into account. In the figures, the y-axis displays how often the participants selected the respective objects (1-8, see Figure 5.1) for each adjective. The raw numbers are given in Table B.4 to B.7 in the Appendix. On the x-axis, the eight objects represented by numbers are plotted; only for the ease of illustration they are ordered linearly (1 = smallest object in RG-trials/dirtiest object in AG-trials, 8 = biggest object in RG-trials/cleanest object in AG trials). Analyzing both trials per adjective together results

[5] A Wilcoxon Signed-Ranks Test revealed that in general the cut-off points did not differ significantly between the two trials (see Table B.1 in the Appendix). Therefore, they were analyzed together. The only exceptions were found in the adult group for *big* (Z = -2.07, p = .039) and in the group of 5-year-olds for *small* (Z = -2.53, p = .011).

in 22 answers per adjective for the 3-year-old group (2 trials x 11 participants), 30 answers per adjective for the 4-year-old group (2 trials x 15 participants), 34 answers per adjective for the 5-year-old group (2 trials x 17 participants) and 52 answers per adjectives for the adults (2 trials x 26 participants).

Figure 5.3 shows that the adults started to consider objects big from object 4 on. Their cut-off points ranged from object 4 to object 8. There was a cut between object 5 and object 6 marking the transition between big and non-big objects. Object 6 was most often selected as the cut-off point. The comparison of the children's and adults' choices in Figure 5.3 indicates that children considered more objects big than adults. In some cases already object 1 was chosen as big. Children's cut-off points ranged from object 1 to object 8. The children's and adults' cut-off points differed significantly, but only in one trial (*big space hoppers*) (Z = -2.92, p = .003). Similar to the adults, objects 5 and 6 constituted the cut-off point most often. This finding is also mirrored in the curves displayed in Figure 5.3, which have a similar shape for children and adults. These results indicate that for both adults and children the cut-off point for *big* is located around the center of the scale.

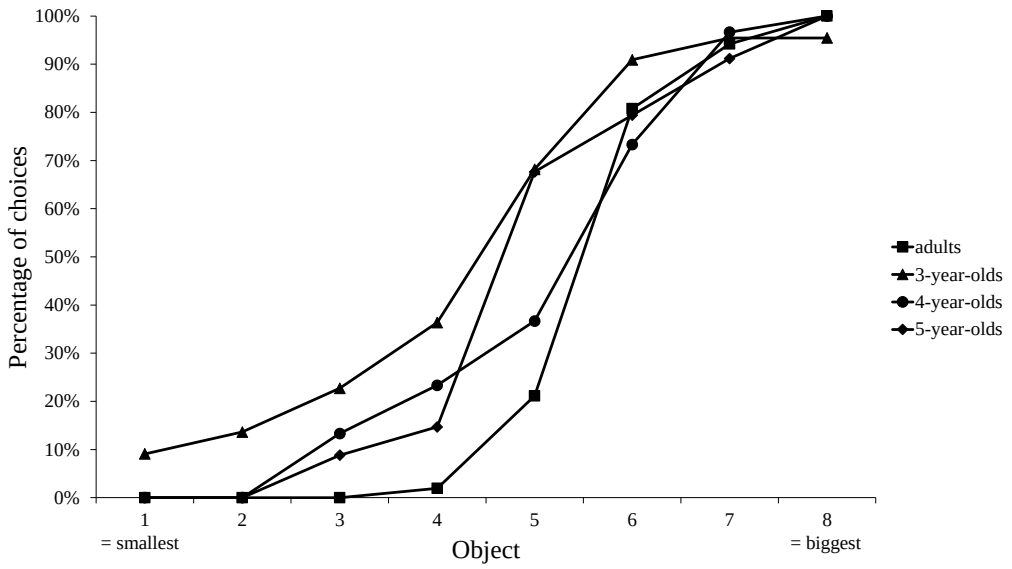


Figure 5.3: Percentage of choices for each object for the RG-trials with *big* per age group.

A closer look into the children's data shows that the interpretation of *big* slightly changed with age. Three-year-old children were more permissive in their judgements for *big*: they selected more often objects at the lower end of the scale than the 4- and 5-year-olds. For the group of 5-year-old children, a clear cut between big and non-big objects could be observed between object 4 and object 5. In contrast, in the group of 4-year-old children the transition between big and non-big objects was less sharp. Overall, no significant difference between the cut-off points of the 3-, 4-, and 5-year-old children was found.

Figure 5.4 illustrates that the adults considered objects small up to object 5. Their cut-off points ranged from object 1 to object 5. Object 3 was most often selected as the

cut-off point. A clear cut could be observed between object 3 and object 4 marking the transition between small and non-small objects. For the children this transition was less sharp, but the children's and adults' cut-off points for *small* did not differ significantly. These findings indicate that for both adults and children the cut-off point for *small* is located around the center of the scale. A comparison of the child groups shows that the 4-year-olds considered more objects small as the 3- and 5-year-olds: in some cases already object 6, the third biggest object, marked the cut-off point. For the 5-year-old group a cut between object 4 and object 5 could be observed marking the transition between small and non-small objects, a pattern similar to the choices for *big*. Overall, all age groups had the same standard of comparison for *small*.

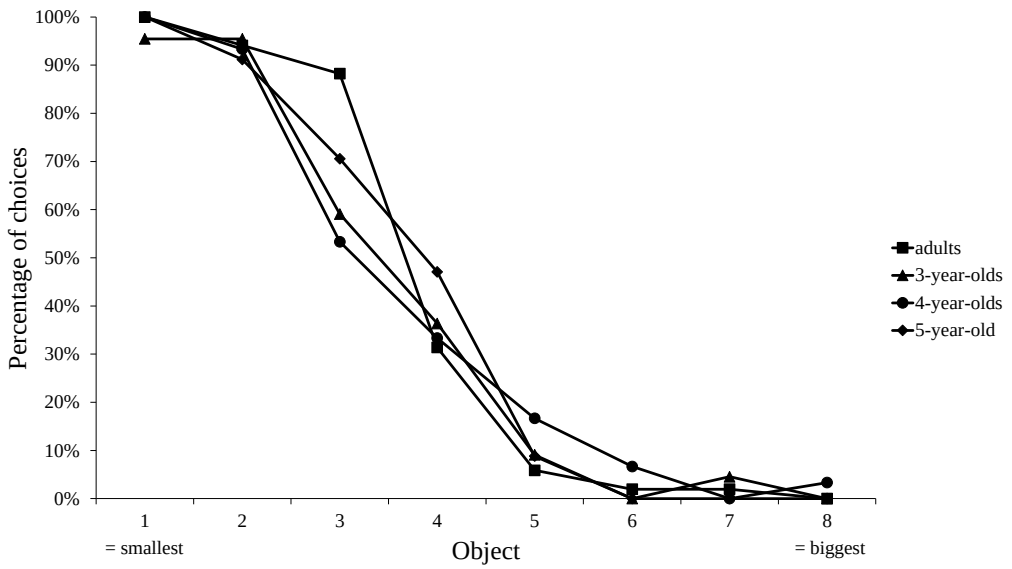


Figure 5.4: Percentage of choices for each object for the RG-trials with *small* per age group.

In sum, the standard for RGs is located around the center of the scale for both children and adults with only small differences between the age groups. For adults, the standard is oriented towards the positive pole of the scale for *big* and towards the negative pole for *small*. The results of the 5-year-old children indicate that they located the standard exactly at the midpoint of the scale. For the 3- and 4-year-old children, the position of the standard was less sharp. The findings suggest that the two younger age groups consider more objects small and big, respectively, than the 5-year-olds and adults.

When asked for the AG *clean* the adults selected only the object that was without any dirt as shown in Figure 5.5. The division of the sequence of objects into clean and non-clean objects was less clear for the children. They sometimes selected also objects with some amount of dirt. But like in the adult group, object 8 was most often selected as the cut-off point. Except for one trial (*clean balls*), for which the difference between the children's and the adults cut-off points was marginally significant (Z = -1.98, p = .048), the cut-off points of the child and the adult group did not differ significantly.

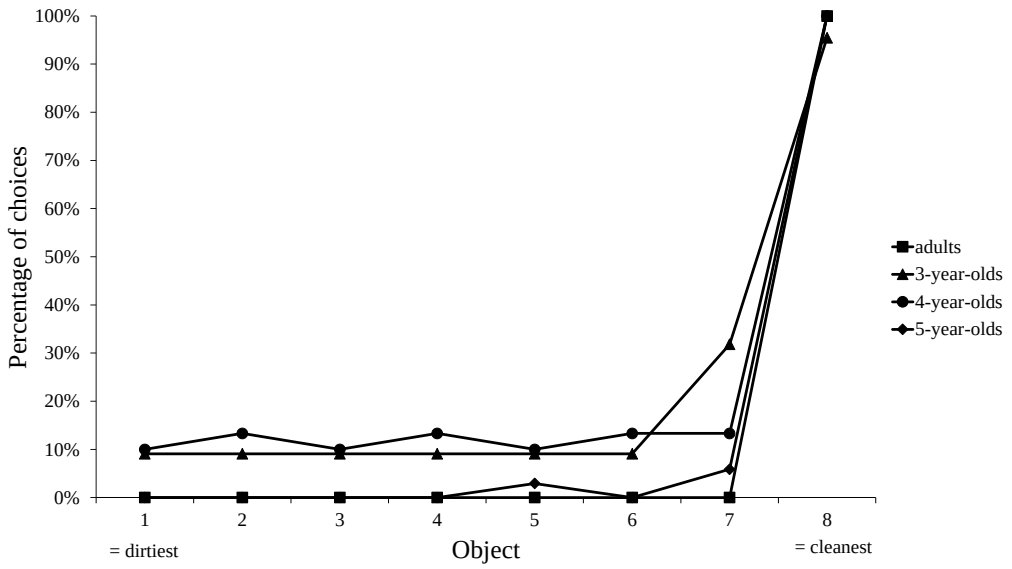


Figure 5.5: Percentage of choices for each object for the AG-trials with *clean* per age group.

The findings indicate that the standard of comparison for *clean* was the maximal degree of the scale for both adults and children. A comparison of the 3-, 4- and 5-year-old group indicates that the two younger age groups differed from the 5-year-old children in that they selected all objects in some of the cases. This difference between age groups is not significant. In all child groups, object 7, which had one spot of dirt on it, was sometimes selected as clean. The number of choices for object 7 decreased with age.

The adults' choices for *dirty* were the mirror image of their choices for *clean* as shown in Figure 5.6. Every object that contained some amount of dirt was selected as dirty. The clean object was never selected. Hence, object 7 marked the cut-off point. The children's choices differed slightly because either not all dirty objects were selected or the clean object was selected as well. However, object 7 was most often chosen as the cut-off point. The children's and adults' cut-off points for *dirty* did not differ significantly.

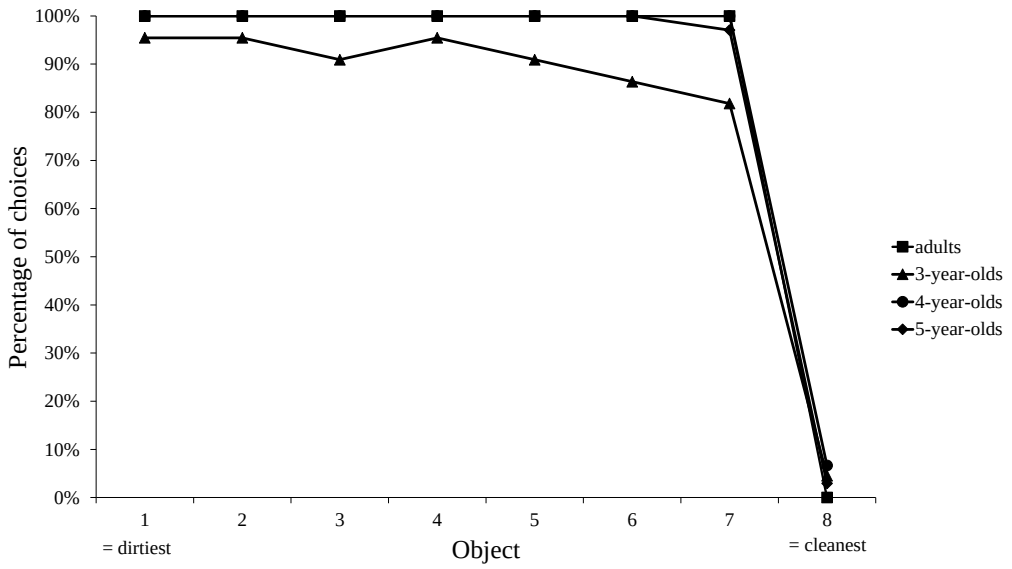


Figure 5.6: Percentage of choices for each object for the AG-trials with *dirty* per age group.

The findings point to a minimal standard of comparison for *dirty* for both adults and children. Compared to the 4- and 5-year-olds, 3-year-old children selected less often every dirty object: objects that were only slightly dirty (e.g., object 6 and 7) were sometimes not considered dirty. Overall, all age groups had the same standard of comparison for *dirty*.

The group data suggest that children distinguish between RGs and AGs when defining the standard as of age 3: for RGs, the standard of comparison is around the center of the scale; for AGs, it is either the maximal or non-zero degree of the scale. However, findings by Tribushinina (2013) showed that 3-year-old children attach the RGs *big* and *small* only to the extremes of the range of objects in the visual context, i.e., only to the smallest or biggest object. Therefore, in a next step individual data were analyzed to explore whether there are participants in the present sample who also used only the extremes of the range of objects as the cut-off point for RGs.

Another individual analysis addresses the question whether antonym pairs are direct opposites and divide the scale exhaustively into for instance big and small objects or whether there are degrees on the scale which count as neither big nor small. This analysis can shed light on the theoretical question whether the standard must be taken to be a single point on the respective scale or a range of degrees. The group results demonstrated that adults oriented the standard for RGs towards one of the poles of the scale, pointing to a gap between the positive and negative antonym. For the group of 5-year-olds the standard for RGs was the exact midpoint of the scale suggesting an exhaustive division of the scale between the positive and the negative antonym. The cut-off points for AGs point to an interpretation of absolute gradable antonyms as direct opposites.

5.4.2.2 Individual data

The first individual analysis examined whether there are participants in the present sample who located the standard of comparison for RGs at the extremes of the scale rather than around the center. Findings by Tribushinina (2013) suggest this response pattern in particular for younger children. The analysis revealed that only few participants (8/69) used the biggest or smallest object as the cut-off point in at least one trial. One 5-year-old used the biggest object as the cut-off point in both *big*-trials, and one 5-year-old used the smallest object as the cut-off point in both *small*-trials. For one 4-year-old the extreme of the object range was the cut-off only in the trials containing space hoppers. Two 4- and 5-year-olds used the biggest/smallest object only in one of the two trials as the cut-off point. For two adults the extreme of the range constituted the cut-off point in every except one trial. For another adult the extreme of the object range occurred only in one trial as the cut-off point. In the group of 3-year-olds, there was no child who attached *big* or *small* exclusively to the extremes of the range of objects.

The second individual analysis investigated whether participants divided the range of objects exhaustively between the negative and positive antonym or whether there were objects which were associated with neither the positive nor the negative adjective. The cut-off points for the negative and the positive adjectives could be compared because the participants saw the identical object arrays once with a test prompt containing the negative antonym and once with a test prompt containing the positive antonym. For RGs, in many of the participants' repsonses a gap between small and big objects was observed. Forty-five of the 69 participants considered some objects consistently, i.e., in all trials, as neither big nor small. Only one adult and three 5-year-old children divided the range exhaustively into small and big objects, i.e., every object that was not considered big was considered small and *vice versa*. This adds to the group results that not all 5-year-old children located the cut-off point exactly in the middle of the scale, and thus have an adult-like interpretation of RGs. For AGs, a gap never existed between clean and dirty objects as already suggested by the group results. Sixty-two of the 69 participants divided the range of objects always exhaustively into clean and dirty objects, i.e., every object that was not considered clean was considered dirty and *vice versa*. The number of participants who either showed or did not show a gap between the standards for negative and posititve RGs and AGs is given in Table 5.2.

Table 5.2: Number of participants who showed a gap or no gap between positive and negative antonym.

		Relative gradable: *big-small*		Absolute gradable: *clean-dirty*	
Age	*N*	Gap	No gap	Gap	No gap
3	11	6	0	0	9
4	15	8	0	0	15
5	17	10	3	0	15
Adults	26	21	1	0	26
Total	69	45	4	0	65

For one 4-year-old, two 5-year-olds and four adults the results in the RG-trials were mixed: a gap between small and big objects existed only in one range of objects (e.g., water balloons), but the other range of objects (e.g., space hoppers) was divided exhaustively into small and big objects. Some children (5/11 3-year-olds, 6/15 4-year-olds, 2/17 5-year-olds) considered some objects both big and small across trials. For three children this was the case for both object arrays, for ten children only for one object array[6]. Four children (2/11 3-year-olds, 2/17 5-year-olds) considered some objects both clean and dirty. As already suggested by the group results, few children (three 3-year-olds, two 4-year-olds, one 5-year-old) selected all objects in some of the trials. I come back to that response type in the next section.

5.4.3 Discussion

The first part of the experiment addressed the questions whether monolingual German-speaking children differentiate between absolute and relative gradable adjectives when defining the standard of comparison (C1a.) and whether the standard of comparison changes with age (C1b.). The findings for German replicate previous findings for other languages (Foppolo & Panzeri, 2013; Syrett, 2007; Syrett et al., 2006) and answer the question of how the standards for positive and negative pole adjectives are related. According to H1, I predicted two outcomes: first, RGs and AGs should be interpreted target-like with respect to the standard of comparison at the same age. Second, if there is a stage at which the distinction between RGs and AGs is not acquired, RGs and AGs should receive the same interpretation. With regard to (C1a.), assuming the semantic analysis in Chapter 2, a target-like response for RGs (*big, small*) was to locate the cut-off point around the center of the scale. For AGs, either the object which exhibits the adjectival property to the maximum (for *clean*) or the minimum (for *dirty*) degree was the target-like cut-off point. Overall, the findings indicate that children as young as age 3 differentiate between RGs and AGs when defining the standard of comparison, hence 3-year-olds interpret RGs and AGs target-like. This result is in line with findings from previous studies on the acquisition of RGs and AGs (Foppolo & Panzeri, 2013; Syrett, 2007; Syrett et al., 2006), which inves-

[6] In the other array, both a gap or no gap were found

tigated children with other first languages and which used a different experimental design. As precited by H1, the findings suggest that RGs and AGs are interpreted target-like with respect to the standard of comparison at the same age.

Although in general children and adults did not differ in their standard of comparison for RGs and AGs, some small differences could be observed in the group data. For RGs, the adults' cut-off point was oriented towards the positive pole of the scale for *big* and towards the negative pole for *small*. In contrast, the children, in particular the group of 5-year-olds, tended to locate the cut-off point in the middle of the scale (see Tribushinina, 2013, for Dutch). For AGs, children's judgements with respect to *clean* and *dirty* were more permissive than the adults' judgements: unlike the adults, they sometimes did not select the minimally dirty objects when asked for *dirty* and they sometimes considered minimally dirty objects clean. Foppolo and Panzeri (2013) and Syrett et al. (2006) found similar responses for *full* in 3- to 5-year-old English- and Italian-speaking children. In Syrett et al.'s (2006) study, one group of children were tested on RGs followed by AGs, the other group with the reverse order. The results show that most of the children who judged also non-maximally full containers as full were tested on RGs first. Hence, children's interpretation of AGs may have been influenced by their interpretation of RGs. In the present study, the first test prompt always contained a relative gradable adjective. This order could be a possible explanation for children's less precise answers compared to the adults. Due to the relatively small set of participants only one order of items was tested. Future studies should counterbalance the order of RGs and AGs to investigate the possible influence of the interpretation of RG-trials on the interpretation of AG-trials.

Addressing (C1b.), the cut-off points chosen by the 3-, 4-, and 5-year-old children did not differ significantly for the respective adjectives. However, small developmental differences exist. With respect to RGs, the group of 3- and 4-year-old children were more permissive because they considered more objects *big* and *small*, respectively, than the 5-year-old group. Regarding AGs, 3- and 4-year-olds differed from the 5-year-olds in that they sometimes selected all objects. This response may have occurred due to linguistic and non-linguistic reasons. A linguistic explanation is given by Ninio (2004). Ninio (2004) investigated the comprehension of adjective-noun combinations in Hebrew-speaking children between 1;6 and 4;4 years. As described in Chapter 3, the children in Ninio's study saw four pictures (e.g., a big and a small clock and a big and a small teddy) and heard a test prompt with an attributive adjective. As the most frequent error children selected a picture that showed the correct object, but the object did not have the property described by the adjective. For instance, when the children were asked *Show me the big teddy*, they pointed to the picture with the small teddy. Ninio (2004) concludes that this error occurs because children have problems with integrating the adjective and noun meaning. As a result, only the noun is processed and the adjective is ignored.

Because 3- and 4-year-olds often selected all objects in *clean*- and *dirty*-trials, a non-linguistic explanation comes into consideration. It may be the case that children did not

recognize the painted spots of dirt as dirt. Hence, for the children the objects did not differ in their properties. However, the non-linguistic explanation seems unlikely for two reasons. First, in the practice trials children saw already some dirty objects and where asked whether they can recognize what the brown spots on the objects could be. If they were not able to recognize the spots as dirt, the experimenter told them that they are made of dirt. But note that crucially the experimenter did not use the word *dirty* to avoid prompting the child for the test trials. Second, children who selected every object did not do so consistently for all *clean*- and *dirty*-items. In addition, selecting all objects occurred also for *big*-trials. Therefore, the linguistic explanation may be favored over the non-linguistic explanations. However, based on the findings from Part 2 and 3 of the experiment presented in Section 5.5 it is unlikely that children are not able to integrate the adjective and the noun meaning. For the time being, the question of why children in some trials selected all objects awaits a conclusive answer.

Interestingly, the present findings do not provide evidence for Tribushinina's (2013) claim that children start with an interpretation of RGs which makes the adjective true only for the object that counts as the best exemplar for the adjective denotation – the biggest or smallest object. The different findings may be ascribed to differences in the visual set-up and in the test prompts. The visual set-up in Tribushinina (2013) consisted of best exemplars of bigness (elephants, hippos, houses, planes), best exemplars of smallness (mice, chickens, gnomes, babies) and four prototype-neutral categories (balloons, cakes, monkeys, umbrellas).[7] As described in Section 5.4.1.1, when testing gradable adjectives it is important that the objects could hold the adjectival property to varying degrees. This is not necessarily the case for the objects used in Tribushinina (2013). As a consequence, children may have focused on the extremes of the object array. The fact that objects were much smaller than in reality could have interfered with children's interpretation of gradable adjectives, in particular for the adjective *big*. For instance, imagine a participant sees pictures of seven mice. Mice are prototypically small and in addition all smaller than in reality because they are presented as pictures. If the participant were asked for *the big mice*, they can be prompted to select only the biggest mouse because this is the one that violates their representation of big mice least. In the present study the objects were chosen such that they do not promote such an interpretation. Moreover, the ordered presentation of objects in Tribushinina (2013) may have prompted the selection of the biggest/smallest object more strongly than the unordered presentation of objects in the present study because the biggest/smallest object may be more prominent in an ordered array of objects than in an unordered one. In addition, the test sentence in Tribushinina's study was a subjective question of the form *Which X do you find {big/small}?*. This way of asking may also have favored answers in which the participants chose only the best exemplar.

Developmental differences were observed with respect to the nature of the standard of RGs. Recall that it is a theoretical question whether the standard should be defined as a

[7] Note that the goal of Tribushinina's (2013) study was to find out whether children are able to include for instance world knowledge in their judgements.

single point, a degree, on the scale or whether it should be defined as a range. Theoretical approaches assume that the former should hold for AGs because negative and positive antonyms are interpreted as direct opposites (Rotstein & Winter, 2004), and that the latter holds for RGs due to the 'gap' between negative and positive antonyms. As both positive and negative antonyms were investigated in the present study, it could be examined how the participants distributed the array of objects. In the AG-trials with *clean* and *dirty*, participants exhaustively classified objects as either clean or dirty. Hence, the results imply that the standard for AGs is defined as a single degree, which is shared by the positive and the negative antonym confirming theoretical approaches by Kennedy (2007) and Rotstein and Winter (2004).

The results of the RG-trials with *big* and *small* revealed that there are objects for which it is difficult to judge whether they are big or not big, or small or not small. This finding is in line with Solt's (2011) account to define the standard for RGs as a range of degrees. In addition, the results suggest that before children arrive at this definition of the standard for RGs, they have at least one prior definition of the standard. The group results of the 3- and 4-year-olds showed that they considered more objects small and more objects big than the 5-year-olds and the adults pointing to an overlap of small and big objects. This is also mirrored in the individual data of the 3-and 4-year-olds. Almost half of them (5/11 3-year-olds, 6/15 4-year-olds) considered some objects both big and small across trials. The other half of the children (6/11 3-year-olds, 8/15 4-year-olds) showed a gap between big and small objects across trials. These response patterns can be explained by two different standards for the positive and the negative antonym. If the standard of the positive adjective is lower than the standard for the negative adjective, some objects are considered both big and small. If the standard of the positive adjective is higher than the standard for the negative adjective, a gap between big and small objects exists. The group results of the 5-year-olds point to an interpretation of the standard for RGs as a single degree because the cut-off point was exactly at the midpoint of the scale for *big* and *small*. However, the individual data revealed that this holds only for some children. More than half of the 5-year-olds (10/17 children) exhibited a gap between big and small objects across trials. Moreover, the individual data revealed stages at which children have two competing definitions of the standard for RGs. There were children that divided the array of objects exhaustively into big and small objects parallel to allow big and small objects to overlap. Other children exhibited a gap between big and small objects parallel to allowing big and small objects to overlap. And a third group of children divided the array of objects exhaustively into big and small objects parallel to exhibiting a gap between big and small objects. Therefore, I propose the following developmental steps for RGs:

1. The standard is one degree exactly at the midpoint of the scale shared by the negative and positive antonym.

2. Negative and positive antonym have different standards.

3. The standard is a range of degrees.

Step 2 can be understood as a precursor of step 3 and can be skipped. This intermediate step mirrors children's knowledge that the standard for RGs is not a single degree. As a result, assuming two separate standards allows for the gap between small and big objects but can lead two responses in which the standards are reversed. This reversal is overcome in the last step.

In summary, the present results suggest an adult-like interpretation for RGs and AGs regarding the standard of comparison from age 3, which results in a distinction between RGs and AGs. As predicted by H1, adjectives of the same complexity are interpreted target-like at the same age. With regard to the nature of the standard, developmental steps were observed for RGs. For AGs, the standard was defined target-like from the outset. Hence, as predicted by H1, there is a stage at which RGs and AGs are interpreted similarly, namely with a standard defined as a single degree. At this point, it seems that in this respect RGs are interpreted initially as AGs. In the next section, it is investigated whether this is also the case with respect to the relevance of the comparison class.

5.5 Part 2 and 3: Relevance of the comparison class

In this section, the second and the third part of the experiment are described and the responses of Part 2 and 3 are compared. This analysis addresses the question whether children are sensitive to changes in the comparison class when calculating the standard of comparison and whether changes in the comparison class affect the standard of all gradable adjectives (C2a.). In addition, it was examined whether this sensitivity changes with age (C2b.). Based on previous research (Barner & Snedeker, 2008; Syrett, 2007; Syrett et al., 2006, 2010) it is expected that 4-year-old children are sensitive to changes in the comparison class for RGs, but probably not for AGs, thus have an adult-like interpretation of AGs and RGs. Because research on 3-year-old's interpretation of gradable adjectives is restricted to implicit comparison classes, it is open how they perform in the present study when the explicit comparison class changes. Findings by Syrett (2007), Syrett et al. (2006), and Syrett et al. (2010) on implicit comparison classes suggest that at age 3, children are able to distinguish between RGs and AGs.

5.5.1 Method

5.5.1.1 Materials

In total, Part 2 and 3 of the experiment consisted of 16 test items and 12 filler items. As in Part 1, the test items included the two relative gradable adjectives *big* and *small* as well as the two absolute gradable adjectives *clean* and *dirty*. In Part 2 and Part 3, the participants saw the same eight objects per trial as in Part 1 (8 water balloons, 8 space hoppers, 8 teddies, 8 balls) and four additional toys from a different basic-level category (4 soccer

balls, 4 buckets, 4 puppets). In Part 2 the adjectives modified the superordinate-level noun *toys*, e.g., *Gib mir bitte die großen Spielsachen* ('Please give me the big toys'), in Part 3 the adjectives modified the same basic-level nouns as in Part 1, e.g., *Gib mir bitte die großen Wasserbomben* ('Please give me the big water balloons'). As shown in the example trials in Figure 5.7, the visual context was identical in Part 2 and Part 3, but the noun in the test prompts differed.

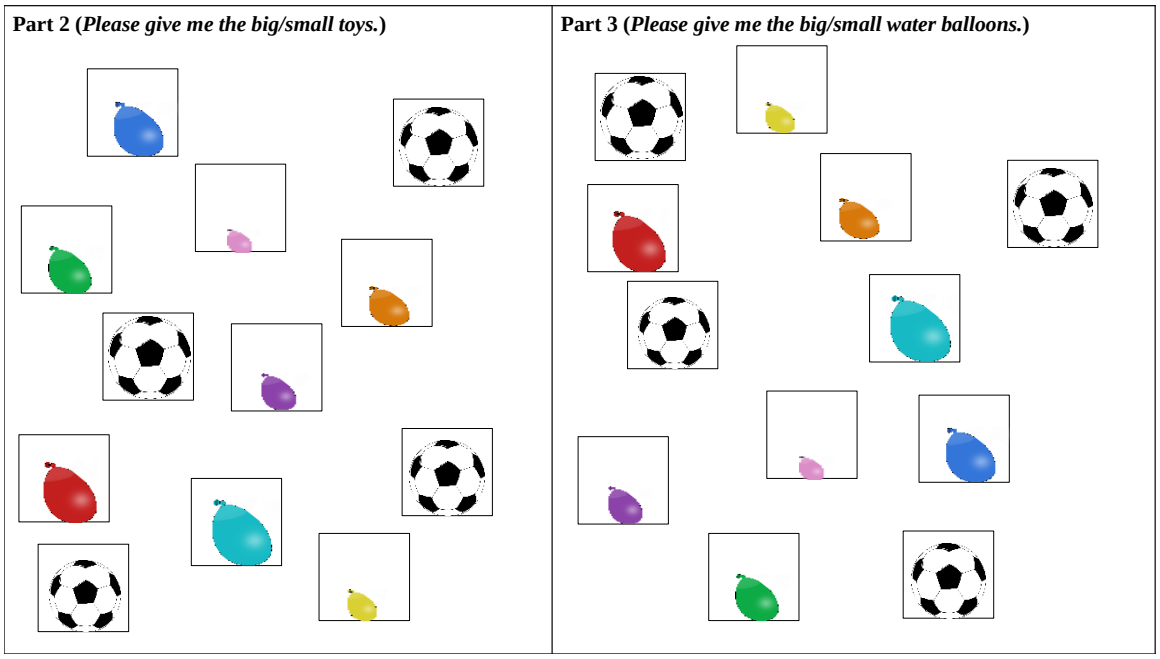


Figure 5.7: Example test trials for the RGs *big* and *small* in the upper expansion context.

By using different linguistic descriptions (superordinate- and basic-level nouns) for the same visual array it was possible to test whether the participants can use the information provided by the noun to determine the comparison class. This is another reason why toys were used as test objects. This way, the same visual array can be described by different nouns. This was achieved by using superordinate- and basic-level nouns. It was also controlled that the nouns were familiar to children as young as age 3. More important, this way it could be tested for which adjectives participants adjust the standard of comparison according to changes in the comparison class. The four additional objects expanded the scale either at the upper end, i.e., they were bigger (as in Figure 5.7) or cleaner, or at the lower end, i.e., they were smaller or dirtier (as in Figure 5.8).

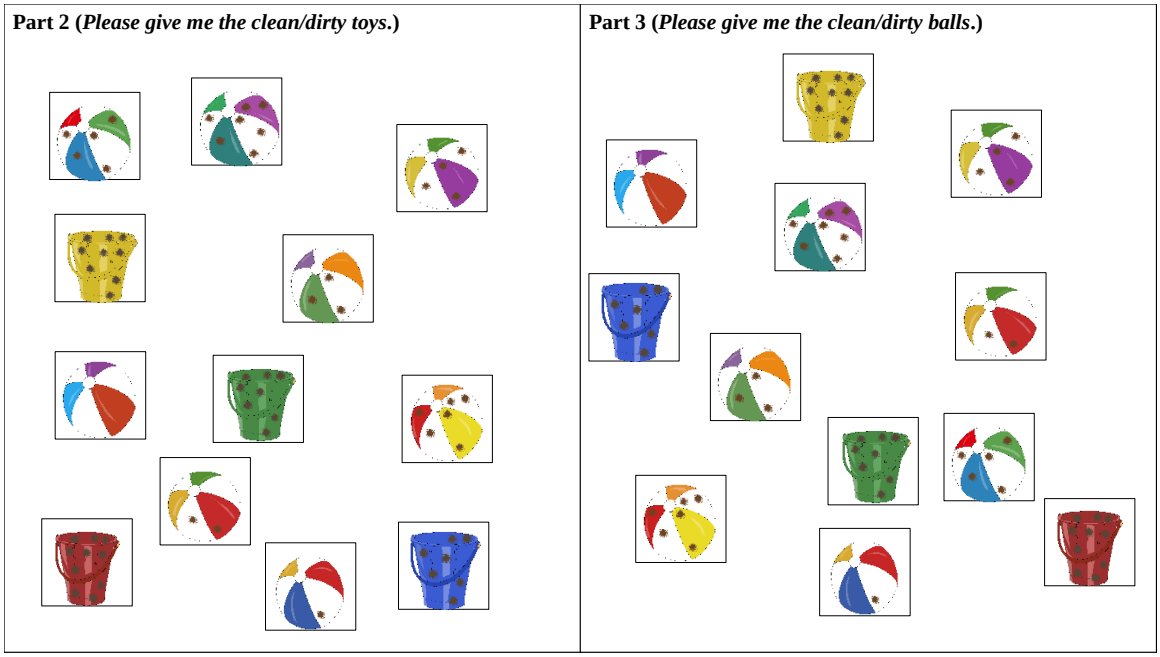


Figure 5.8: Example test trials for the AGs *clean* and *dirty* in the lower expansion context.

The expansion of the scale at the upper or the lower end assessed whether participants are sensitive to the statistical properties of the object array. Adding objects at the upper end, the so-called 'upper expansion context', increases the average amount of the adjectival property, e.g., the mean height for toys. Adding objects at the lower end, the so-called 'lower expansion context', reduces the average amount of the adjectival property. If participants adjust the standard to the comparison class encoded by the noun and consider the statistical properties of the object array, they should shift the standard in different directions depending on the type of expansion. Consider for example the test trial in Figure 5.7. When asked for *big water balloons* (Part 3) the water balloons serve as the comparison class. When asked for *big toys* (Part 2) the soccer balls should be included in the comparison class as well. Because they are bigger than the water balloons, the mean size of the comparison class increases. Therefore, the standard for *big toys* should be higher than the standard for *big water balloons*. The reverse holds for visual contexts with additional objects that are smaller than the water balloons.

As in Part 1, the same visual contexts were used for *big* and *small* on the one hand, and for *clean* and *dirty* on the other hand. Thus, it was directly accessible whether children's responses differed between positive and negative gradable adjectives. As described in Chapter 3, previous findings (e.g., Barner & Snedeker, 2008) indicate that children interpret positive adjectives such as *tall* more often target-like than negative adjectives such as *short*. Table 5.3 summarizes the materials of Part 2 and 3.

The filler items had the same structure as the filler items in Part 1. All test and filler items and the corresponding pictures are displayed in Appendix B.4.

Table 5.3: Test trials.

Adjective	Modified noun	Visual Context[a]	
		upper expansion	lower expansion
RG (*big, small*)	basic-level	8 water balloons + 4 bigger soccer balls	8 space hoppers + 4 smaller buckets
	superordinate-level	8 water balloons + 4 bigger soccer balls	8 space hoppers + 4 smaller buckets
AG (*clean, dirty*)	basic-level	8 teddies + 4 cleaner dolls	8 balls + 4 dirtier buckets
	superordinate-level	8 teddies + 4 cleaner dolls	8 balls + 4 dirtier buckets

[a] Except for the soccer balls, each object was depicted in a different color.

5.5.1.2 Procedure

As in Part 1, the positive RG and AG (*big, clean*) were tested in the first session, and the negative counterparts (*small, dirty*) were tested in the second session. After the unrelated task following Part 1 participants continued with Part 2 and Part 3. Both sessions started with a trial containing a relative gradable adjective and then alternating absolute and relative gradable adjectives. For each adjective, the trial with the upper expansion context was presented before the trial with the lower expansion context. As in Part 1, test items and filler items were intermixed resulting in two test items followed by two filler items etc. The task was identical to Part 1 except that more picture cards were presented and the noun in the test prompt differed.

5.5.1.3 Data analysis

Regarding the question which gradable adjectives participants interpret relative to a comparison class, it was coded whether participants had different cut-off points for the trials in Part 2 compared to the trials in Part 3. Changes in the comparison class were encoded by different nouns (superordinate- vs. basic-level) in the test prompts describing the identical visual array. This way, it was possible to examine whether participants used the information provided by the noun to determine the comparison class. More important, it was analyzed for which adjectives changes in the comparison class resulted in an adjustment of the cut-off point, i.e., the standard of comparison.

Following semantic theory, for AGs, a response was coded as target-like if the cut-off point was the same for different nouns. For RGs, a response was coded as target-like if the cut-off point differed for different nouns, i.e., was adjusted to changes of the noun. In addition, sensitivity to the statistical properties of the visual array was investigated. This means a response for RGs was only coded as target-like if the standard was shifted in the correct direction given the distribution of objects in the visual array. As explained above, in the upper expansion context, the standard for *big/small toys* should be "higher" than for *big/small water balloons*. In contrast, the standard for *big/small toys* should be "lower" than

the standard for *big/small* space hoppers in the lower expansion context. This is illustrated in Figure 5.9 for *big*. In this hypothetical scenario for the upper expansion context, the participant's cut off-point for *big water balloons* is the circled blue water balloon on the right picture (= object 6), whereas for *big toys* the cut-off point is the circled turquoise water balloon on the left picture (= object 8).

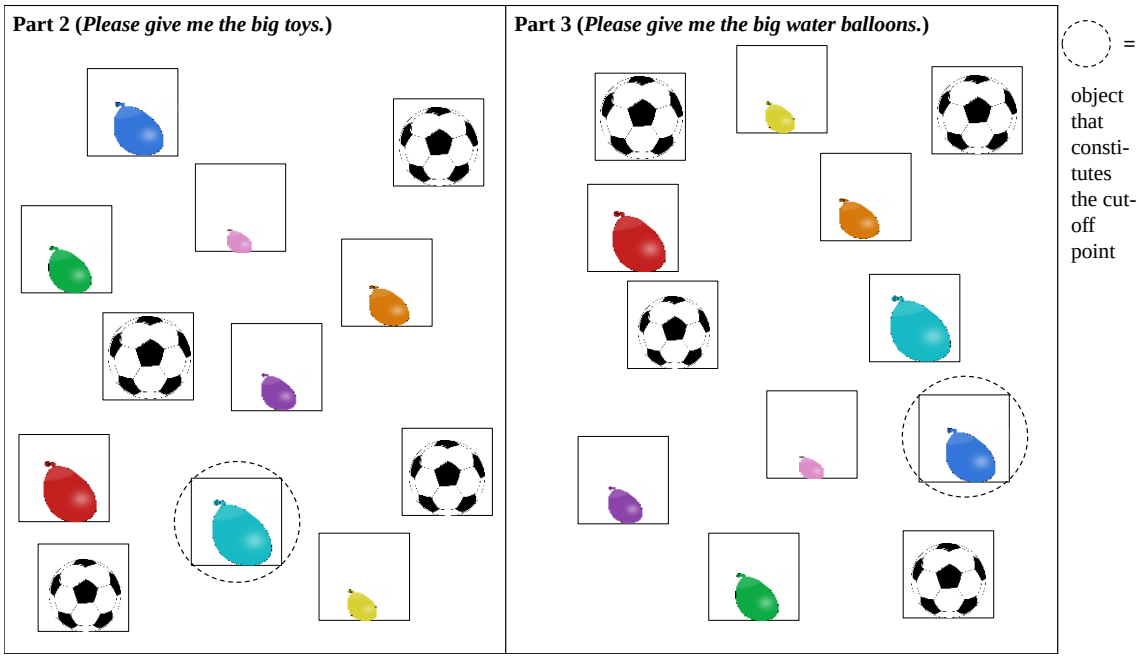


Figure 5.9: Example shift of the cut-off point for the RG *big* in the upper expansion context.

Similar to Part 1, responses were coded as unanalyzable if not all objects following the cut-off point were selected, e.g., if the blue and turquoise water balloons (= objects 6 and 8) were selected when asked for *big water balloons*, but the red water balloon (= object 7) was not.

5.5.2 Results

5.5.2.1 Group data

In Part 2 and 3 of the experiment, each adjective occurred once in the upper expansion context and once in the lower expansion context. In Part 2, the adjectives always modified the superordinate-level noun *toys*. In Part 3, the participants saw the same visual stimuli as in Part 2, but the adjectives modified the same basic-level nouns as in Part 1. It was analyzed for which items the participants adjusted the standard of comparison according to the noun. This means the participants had the chance to adjust the standard four times for RGs (twice for *big* = once in upper expansion context, once in lower expansion context; twice for *small* = once in upper expansion context, once in lower expansion context) and four times for AGs (twice for *clean* = once in upper expansion context, once in lower expansion context; twice for *dirty* = once in upper expansion context, once in lower expansion

context). Following semantic theory, an adjustment of the standard was expected for RGs, but not for AGs. It was analyzed how often the participants adjusted the standard across trials for RGs and for AGs. For RGs, the adult group adjusted the standard according to the noun in 45.2% of the anwers. For AGs, they adjusted the standard in 1.0% of the answers. Adults adjusted the standard significantly more often according to the noun for RGs than for AGs (Z = -4.28, $p < .001$). This significant difference in standard adjustments between RGs and AGs was also found in the child group. In 36.6% of the answers the standard was adjusted for RGs and in 8.7% of the answers for AGs (Z = -4.44, $p < .001$). These data indicate that children interpret RGs relative to a comparison class, but not AGs.

To find out whether interpretation patterns change with age, each age group was analyzed individually with regard to the number of standard adjustments. Table 5.4 summarizes how often participants adjusted the cut-off point according to the noun across trials. Raw numbers are given in Table B.8 in the Appendix.

Table 5.4: Percentage of standard adjustments per age group and adjective class.

Age	N	Relative gradable: *big/small* (n=4)	Absolute gradable: *clean/dirty* (n=4)
3	11	22.7	18.2
4	15	40.0	3.3
5	17	42.7	7.4
Adults	26	45.2	1.0

The group of 3-year-old children adjusted the standard of comparison equally often for relative and absolute gradable adjectives (Z = -.541, p = .589). In contrast, 4- and 5-year-old children adjusted the standard of comparison significantly more often for relative than for absolute gradable adjectives (age 4: Z = -3.11, p = .002; age 5: Z = -3.37, p = .001).

Three questions follow from the above findings:

1. Can the participants' answers be ascribed to specific adjectives or specific visual contexts?

2. Do 3-year-old children interpret RGs as AGs or *vice versa*?

3. Why do adults adjust the standard for RGs only in about half of the cases?

The first question relates to previous findings which suggest that children interpret positive adjectives more often target-like than negative adjectives. Data from German can reveal whether the advantage of positive over negative adjective holds cross-linguistically or whether the previous findings can be attributed to specific properties of English. Barner and Snedeker (2008) note that the children in their study insisted that the shortest object

in the array was not *short* but rather *small* or *little*. According to Barner and Snedeker (2008, p. 606), this response can be explained by a "mutually exclusive interpretation of *short* and *small* (where *small* applies to the shortest objects and *short* to middling values)". This mutual exclusivity does not exist in German. In addition, the experiments that are comparable with the present design investigated only visual contexts in which the scale was expanded at the lower end (see Barner & Snedeker, 2008, Experiment 3 and 4). Hence, it is open whether the visual context influences the interpretation of gradable adjectives, in particular RGs.

The second question sheds light on children's initial semantic representation of gradable adjectives, which are a case in point for adjective classes of equal complexity. The third question is addressed in Section 5.5.2.2, which focuses on the participants' individual data. The first and the second question are addressed in the following.

The number of standard adjustments for *big* and *small* across both upper and lower expansion contexts are displayed in Figure 5.10. The age groups were analyzed individually to detect possible developmental changes. The number of standard adjustments was higher for *big* than for *small* across age groups. This difference did not reach significance for any of the age groups (age 3: Z = -.63, p = .527; age 4: Z = -.97, p = .33; age 5: Z = -.83, p = .405; adults: Z = -.63, p = .527).

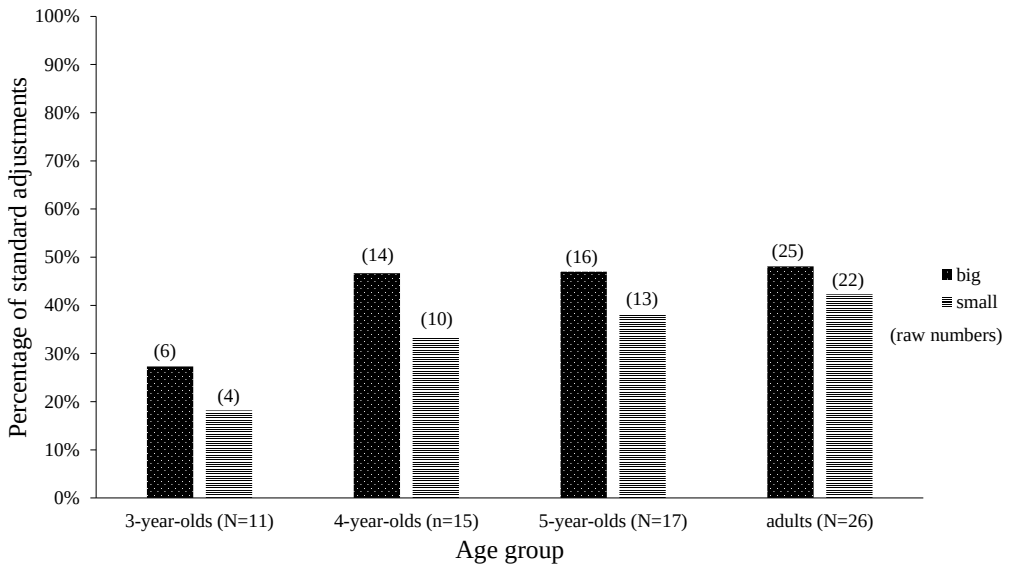


Figure 5.10: Percentage of standard adjustments per age group and relative gradable adjective. 100% equates with the number of participants multiplied by the number of trials (= 2 per adjective).

The number of standard adjustments for *clean* and *dirty* across both upper and lower expansion contexts are displayed in Figure 5.11. If anything, the group of 4-year-olds adjusted the standard only in the *dirty*-trials, the adults only in the *clean*-trials, and the 5-year-olds more often in the *clean*- than in the *dirty*-trials. These differences did not yield significance for any of the age groups (age 4: Z = -1.41, p = .157; age 5: Z = -1.34, p =

.180; adults: Z = -1.00, p = .317). The group of 3-year-olds adjusted the standard equally often for *clean* and *dirty*.

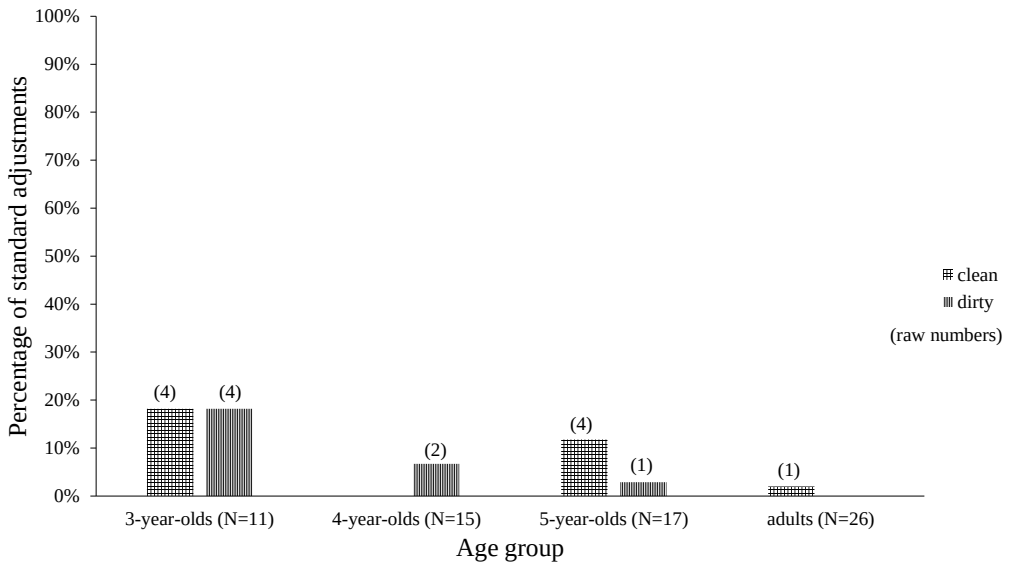


Figure 5.11: Percentage of standard adjustments per age group and absolute gradable adjective. 100% equates with the number of participants multiplied by the number of trials (= 2 per adjective).

In summary, for RGs, the number of standard adjustments was slightly higher for the positive adjective *big* than for the negative adjective *small* in the child and in the adult group. For AGs, the number of standard adjustments was generally very low from age 4 (see Table 5.4). In the group of 3-year-olds, which adjusted the standard more often for AGs than the 4- and 5-year-olds and the adults, the adjectives' polarity did not affect the number of standard adjustments.

A closer look at the two *big*- and the two *small*-items, respectively, examined the influence of the visual context on the interpretation of *big* and *small*. The data in Figure 5.12 indicate that the number of standard adjustments for *big* was higher in upper expansion contexts than in lower expansion contexts across age groups. The difference was significant only in the group of 5-year-old children (Z = -2.83, p = .005). The difference between the upper and the lower expansion context was neither significant for 3-year-olds (Z = -.82, p = .414), nor for 4-year-olds (Z = -.82, p = .414), nor for the adults (Z = -.28, p = .782).

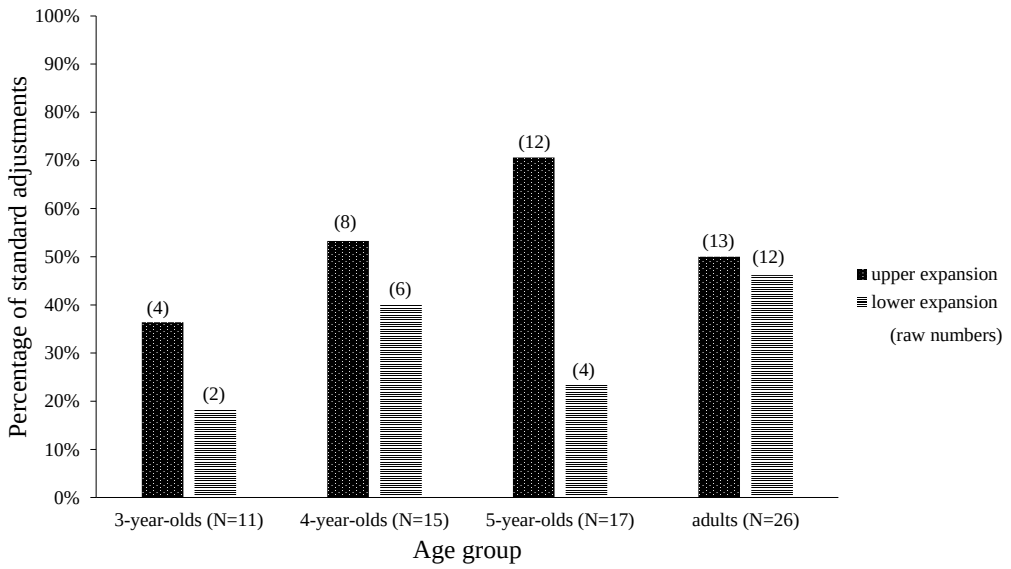


Figure 5.12: Percentage of standard adjustments for the RG *big* per visual context and age group. 100% equates with the number of participants multiplied with the number of trials (= 1 per visual context).

The data in Figure 5.13 demonstrate that the number of standard adjustments for *small* was slightly higher in the lower expansion context than in the upper expansion context. The difference did not reach significance for any of the age groups (age 3: Z = -1.41, p = .157; age 4: Z = -.82, p = .414; age 5: Z = -1.00, p = .317; adults: Z = -1.60, p = .109).

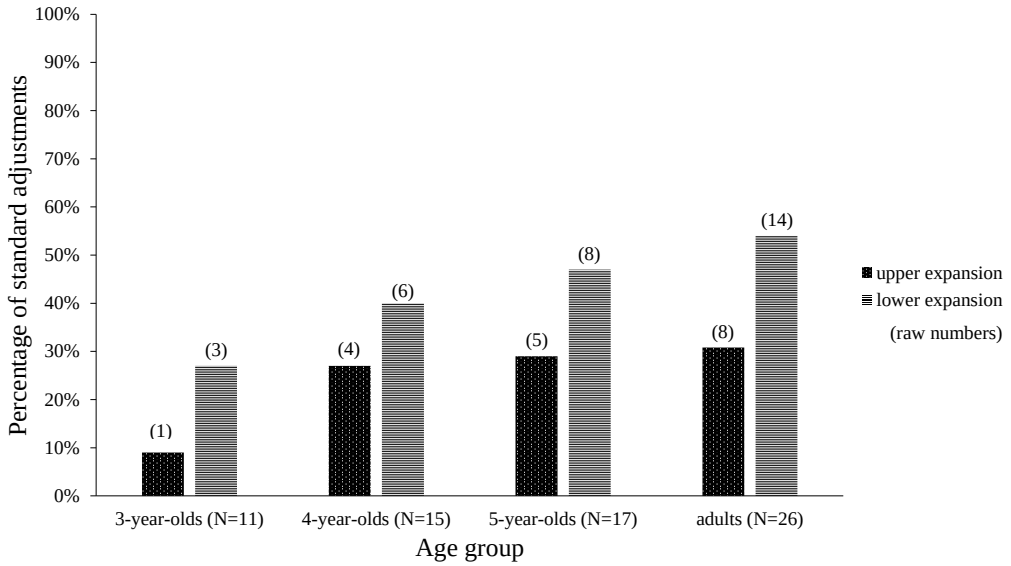


Figure 5.13: Percentage of standard adjustments for the RG *small* per visual context and age group. 100% equates with the number of participants multiplied with the number of trials (= 1 per visual context).

Regarding the number of standard adjustments for RGs, the data in Figures 5.12 and 5.13 suggest a relation between the adjectives' polarity and the kind of expansion. In the

case of *clean* and *dirty*, respectively, the data of the 3-year-old group are most informative regarding the influence of the visual context because the 4- and 5-year-old and the adult group hardly ever adjusted the standard for AGs. The data in Figure 5.14 and 5.15 suggest that the relation between the adjective's polarity and the kind of expansion holds for AGs as well: the standard for the positive adjective *clean* was slightly more often adjusted in upper expansion contexts. In contrast, the standard for the negative adjective *dirty* was slightly more often adjusted in lower expansion contexts. The difference between upper and lower extension contexts did not reach significance (age 3: Z = -1.00, p = .317 and Z = -1.41, p = .157).

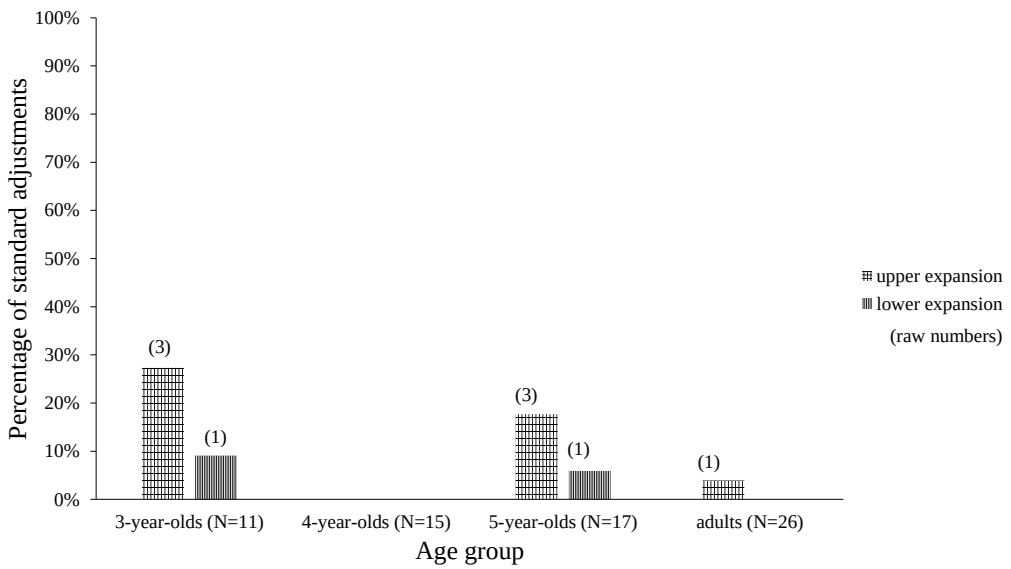


Figure 5.14: Percentage of standard adjustments for the AG *clean* per visual context and age group. 100% equates with the number of participants multiplied with the number of trials (= 1 per visual context).

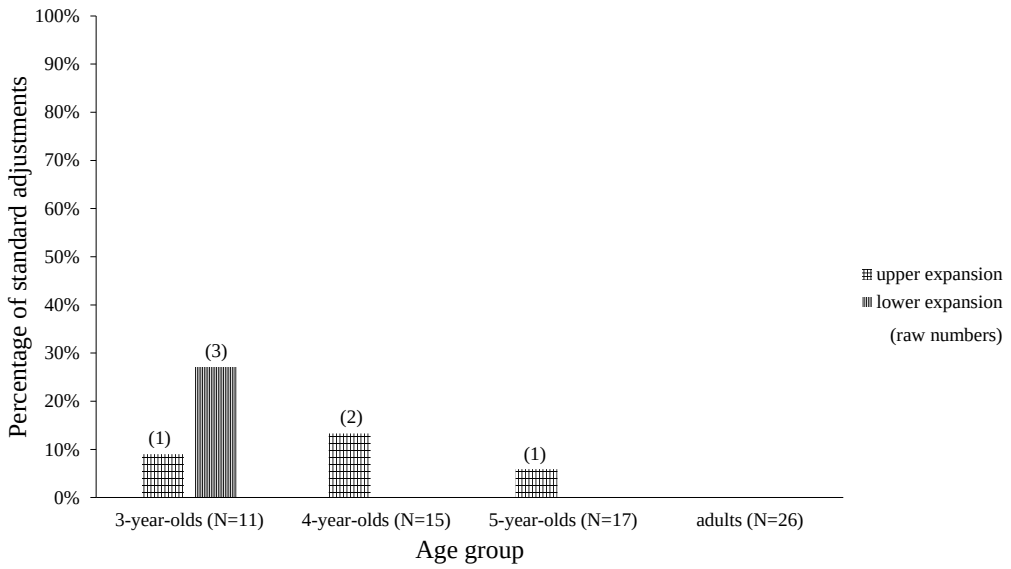


Figure 5.15: Percentage of standard adjustments for the AG *dirty* per visual context and age group. 100% equates with the number of participants multiplied with the number of trials (= 1 per visual context).

With regard to the first question whether the participants' answers can be ascribed to specific adjectives or specific visual contexts, two trends were identified. First, particularly for RGs the standard for the positive adjective *big* was preferably adjusted compared to the negative adjective *small*. Second, the number of standard adjustments was higher when the adjective's polarity coincided with the pole of expansion.

The results displayed in Table 5.4 show that the group of 3-year-olds adjusted the standard of comparison equally often for RGs (22.7%) and for AGs (18.2%). A comparison of the number of standard adjustments in the 3-year-old group with the adult group yielded significant differences for both RGs and AGs. As the numbers in Table 5.4 suggest, the 3-year-olds adjusted the standard less often than the adults for RGs (Z = -2.55, p = .011). For AGs, the 3-year-olds adjusted the standard more often than the adults (Z = -3.57, $p < .001$). Hence, it seems that 3-year-old children interpret neither RGs nor AGs adult-like with respect to the relevance of the comparison class. Therefore, it was analyzed what children's specific standard adjustments looked like. If the 3-year-olds adjusted the standard for RGs, they adjusted it in an adult-like way. The 3-year-olds' standard adjustments for AGs looked different. There is only one case which is comparable to the standard adjustments for RGs: in the positive expansion context, the child selected only the maximally clean objects when asked for *clean toys* but when asked for *clean teddies*, she selected the teddy with one spot of dirt in addition to the maximally clean teddy. The other standard adjustments were such that the children selected the expected objects in Part 2 or Part 3, but selected all toys or all balls/teddies in the respective other part. Because these answers show no sytematic pattern, it is difficult to draw any conlcusion about the 3-year-olds' interpretation. When asked for *clean toys*, two children did not select the clean teddy but only the clean dolls in the positive expansion context. This is

not a real standard adjustment because the dolls and the clean teddy were equally clean. Although the standard adjustments are not identical for RGs and AGs, the comparison class influences the 3-year-olds' behavior with respect to both in a non-adult-like way.

In summary, the group results of Part 2 and 3 of the experiment suggest that as of age 4, children interpret RGs, but not AGs, relative to a comparison class. At age 3, the distinction between RGs and AGs with respect to the relevance of the comparison class is not yet adult-like. The next section takes individual data into account to identify possible "strategies" under which circumstances the participants adjusted the standard of comparison according to the noun. Moreover, individual data can reveal response patterns that deviated from the expected answer.

5.5.2.2 Individual data

This section aims at showing whether the group data are mirrored in the individual data. In addition, individual data can help to figure out "strategies" in which contexts participants adjusted the standard. First, the data for RGs are reported and then the data for AGs.

In total, only seven of the 69 participants never adjusted the standard for RGs. This means that the majority of participants (8/11 3-year-olds, 14/15 4-year-olds, 17/17 5-year-olds, 23/26 adults) adjusted the standard at least once for RGs. Therefore, the group data cannot be reduced to specific participants. Instead, most participants adjusted the standard for RGs, but not consistently. Thus, it is likely that participants adjusted the standard only under specific circumstances. The group data already suggested a trend regarding a preference for adjusting the standard for *big* when the expansion of the scale matches the polarity of the adjective, e.g., when the scale was expanded at the upper end and the participants were asked for *big toys*. Five response patterns could be observed in the individual data: a) the standard was never adjusted, b) the standard was always adjusted[8], c) the standard was adjusted for one adjective, e.g., *big*, d) the standard was adjusted for one visual context, e.g., in upper expansion contexts, and e) the standard was adjusted in particular combinations of adjective and visual context, e.g., for *big* in upper expansion contexts. The response patterns are summarized in Table 5.5. No systematic pattern could be observed. This holds in particular for the two younger groups.

Overall, the most prominent patterns were b) and two instances of e). Thirteen participants (almost) always adjusted the standard, 8 participants adjusted the standard for *big* in upper expansion contexts as well as for *small* in lower expansion contexts, and 10 participants only for *big* in upper expansion contexts. Across age groups, most participants belonged to group e): they adjusted the standard only in particular combinations of adjective and visual context. Hence, the individual data confirm the trend observed in the group data. The question of how the participants responded in trials in which they did not adjust the standard is addressed in Section 5.5.2.3.

For AGs, the majority of participants (5/11 3-year-olds, 13/15 4-year-olds, 12/17 5-

[8] Participants who adjusted the standard with one exception also fall in this category.

Table 5.5: Number of participants per individual response pattern for relative gradable adjectives: standard adjustments.

Response pattern	Age				Total
	3	4	5	Adults	
a) Never	0	0	0	3	3
b) Always	0	3	4	6	13
c) One adjective					
big	0	2	1	2	5
small	1	0	0	1	2
d) One visual context					
Upper expansion	0	0	1	0	1
Lower expansion	0	0	0	4	4
e) Combination of adjective and visual context					
big-upper expansion + *small*-lower expansion	1	1	3	3	8
big-lower expansion + *small*-upper expansion	0	0	0	1	1
big-upper expansion	2	3	3	2	10
big-lower expansion	1	1	0	0	2
small-upper expansion	0	0	1	2	3
small-lower expansion	1	1	2	1	5
f) No strategy	5	4	2	1	12
Sum	**11**	**15**	**17**	**26**	**69**

year-olds, 25/26 adults) did never adjust the standard. The 14 participants who adjusted the standard did it only once (except for one 3-year-old child) without a systematic pattern. The children who adjusted the standard for AGs were not necessarily the same children who adjusted the standard for RGs, hence I do not assume that these children have transferred their interpretation for RGs to AGs.

5.5.2.3 Alternative answers to RG-trials

The results in Table 5.4 suggest that even the adult participants adjusted the standard only in about half of their responses in RG-trials. However, the individual data presented in Section 5.5.2.2 showed that only three adults and no child never adjusted the standard. Hence, participants must have deployed alternative responses.

Unanalyzable answers

First of all, unanalyzable answers occurred. Responses were coded as unanalyzable if not all objects following the cut-off point were selected, e.g., if objects 6 and 8 were selected, but object 7 was not. If this happened, it could not be determined whether the standard differed between the items with superordinate-level and basic-level nouns. In the group of 3-year-old children, it was not possible to analyze whether the cut-off point was shifted in 37.5% of all answers, in the group of 4-year-old children in 15% of all answers, in the group of 5-year-old children in 8.1% of all answers and in the adults in 1.9% of all answers.

The raw numbers in Table 5.6 reveal that the number of unanalyzable answers was slightly higher in relative than in absolute gradable adjectives. In the class of relative gradable adjectives the answers for the negative adjective *small* were more often unanalyzable than for the positive adjective *big*. The number of unanalyzable answers decreased with age.

Table 5.6: Raw numbers of unanalyzable answers per age group and adjective class.

Age	Absolute gradable		Relative gradable		Total
	clean	*dirty*	*big*	*small*	
3	7	6	7	13	33
4	5	3	3	7	18
5	2	3	1	5	11
Adults	0	0	1	3	4
Sum	**14**	**12**	**12**	**28**	**66**

Standard adjustment in the unexpected direction

Besides the unanalyzable answers, another alternative response was to adjust the standard between the superordinate-level and the basic-level noun in the unexpected direction. Due to the statistical properties of the visual array it was expected that in the upper expansion context, the standard for *big/small toys* should be "higher" than for *big/small water balloons*. In contrast, the standard for *big/small toys* should be "lower" than the standard for *big/small space hoppers* in the lower expansion context. For some participants it was the opposite, though. This response occurred in 18.2% (8/44) of the answers for RGs in the group of 3-year-olds, in 26.7% (16/60) of the answers in the group of 4-year-olds, in 8.8% (6/68) of the answers in the group of 5-year-olds, and in 10.6% (11/104) of the answers in the adult group. This response mostly occurred for single trials only and it was not restricted to specific visual arrays.

Two different standards for each basic-level category

Another alternative response was the calculation of two different standards for each basic-level category in the visual context, when the test prompt contained the superordinate-level noun *toys*. For instance, when asked for *the big toys*, participants calculated one standard for water balloons and another standard for soccer balls, hence they selected the biggest of the water balloons and the biggest of the soccer balls. This response occurred in 11.4% (5/44) of the answers in the group of 3-year-olds, in 25.0% (15/60) of the answers in the group of 4-year-olds, in 11.8% (8/68) of the answers in the group of 5-year-olds, and in 23.1% (24/104) of the answers in the adult group. An individual analysis revealed that only 3 participants (two 4-year-old children and one adult) always calculated separate standards. Thirty-six of the 69 participants (7 3-year-olds, 6 4-year-olds, 11 5-year-olds, 12 adults) never calculated separate standards. One may speculate that the participants

used two separate standards only and always if the standard was not adjusted. However, this was the case only for 5 of the 33 participants who calculated two separate standards at least once. Thus, there were participants who had two separate standards and additionally adjusted the standard. Therefore, it is unlikely that the "strategy" of calculating two separate standards was used to avoid shifting the standard. As can be seen in Table 5.7, it is rather the visual context (in combination with specific adjectives) that triggered the calculation of two separate standards for the two basic-level categories subsumed under the superordinate-level noun *toys*.

Table 5.7: Number of participants per individual response pattern for relative gradable adjectives: calculation of two separate standards for each basic-level category in superordinate-level noun contexts.

Response pattern	Age				Total
	3	4	5	Adults	
a) Never	7	6	11	12	36
b) Always	0	2	0	1	3
c) One adjective					
big	1	0	0	0	1
small	0	0	0	2	2
d) One visual context					
Upper expansion	0	1	1	2	4
Lower expansion	0	0	1	0	1
e) Combination of adjective and visual context					
big-upper expansion + *small*-lower expansion	0	0	0	2	2
big-lower expansion + *small*-upper expansion	0	1	0	0	0
big-upper expansion	0	0	1	4	5
big-lower expansion	0	0	1	0	1
small-upper expansion	1	2	1	0	4
small-lower expansion	2	3	1	3	9
Sum	**11**	**15**	**17**	**26**	**69**

In the next section I discuss how the results of Part 2 and Part 3 relate to the research questions (C2a.) and (C2b.), and how the findings can be explained.

5.5.3 Discussion

The effect of the comparison class on the interpretation of gradable adjectives: standard adjustments

Part 2 and 3 of the experiment addressed the question which gradable adjectives children interpret relative to a comparison class (C2a.). This question was explored by investigating children's sensitivity to changes in the explicit comparison class. In other words, it was analyzed for which adjectives children change their standard of comparison according to changes of the noun in the test prompt. Crucially, the visual array remained the same. Hence, children had to use the information provided by the noun to determine the com-

parison class rather than the visual context. According to semantic theory, the comparison class is relevant for calculating the standard of comparison for RGs, but not for AGs. As a consequence, the standard of comparison for RGs, but not for AGs, should be adjusted according to changes of the noun, i.e., the explicit comparison class. The findings indicate that 4- and 5-year-old children interpret RGs, but not AGs, relative to the explicit comparison class. Thus, as of age 4, children have an adult-like interpretation of gradable adjectives regarding the relevance of the comparison class because they distinguish RGs and AGs. This is in line with hypothesis H1, which predicted that adjectives of the same complexity should be interpreted target-like at the same age. The present findings extend Barner and Snedeker's (2008) findings regarding 4-year-old English-speaking children's interpretation of RGs.

Two observations, in particular with regard to the responses for RG-trials, merit a closer look. First, participants showed a trend to adjust the standard more often for the positive adjective *big* than for the negative adjective *small*. Second, participants showed a tendency to adjust the standard more often when the polarity of the adjective coincided with the polarity of the visual expansion of the object array. For instance, participants were more likely to adjust the standard for *big* according to the noun (*water balloons* vs. *toys*) in upper expansion contexts (see Figure 5.7). In such a context, the four additional toys (soccer balls) were bigger than the eight original toys (water balloons).

An asymmetry between negative and positive adjectives has also been reported in previous studies (e.g., Barner & Snedeker, 2008). Barner and Snedeker (2008) note that the source of the asymmetry is unclear. In their study, the visual context was similar to the lower expansion context of the present study, i.e., the additional objects were always smaller than the original objects. Barner and Snedeker (2008) assume that children failed to take these objects at the lower end of the scale into account when they calculate the standard because these objects lack height. As a consequence, these objects do not cause a change in the standard of comparison. However, in the present study, the standard for *small* was more often adjusted in this lower expansion context. Because the asymmetry between negative and positive relative gradable adjectives was not significant in the present data, more children should be tested to find out whether the difference disappears or whether it becomes significant. In the latter case, an explanation for this asymmetry should be tackled.

The visual context also plays a role for the second trend observed in the data. The findings suggest a preference to adjust the standard in particular combinations of specific adjectives and specific visual contexts. For *big* the standard was preferably adjusted in the upper than in the lower expansion context, for *small* it was the opposite. These data relate to previous findings on adults' interpretation of RGs (Hansen & Chemla, 2017; Ryalls & Smith, 2000). Ryalls and Smith (2000) report on the so-called 'semantic congruity effect'. This effect describes that participants show faster reaction times when they have to select for instance the bigger of two big items than when they have to select the smaller of two big items. This effect has also been found regarding the accuracy of comparative judgements

and the ease of learning for novel relative gradable adjectives. This semantic congruity effect is also mirrored in Hansen and Chemla's (2017) findings. Hansen and Chemla (2017) used the Presupposition Assessment Task by Syrett (2007) and Syrett et al. (2006) with adults. In this task, participants were presented with two objects and were asked to select *the tall one* for instance. Due to the context-sensitivity of RGs participants should always select one object because in a given context one object is always the tall one relative to the other object. Hansen and Chemla (2017) used aliens of four different sizes which they labeled as 0/3, 1/3, 2/3 and 3/3[9] as displayed in Figure 5.16. Similar to Barner and Snedeker (2008), they used objects for which the evaluation of the property cannot be based on world knowledge.

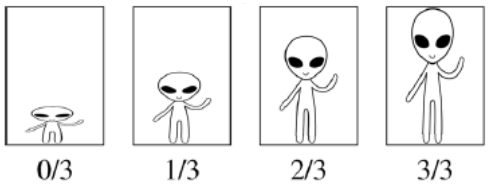


Figure 5.16: Example visual stimuli used in Hansen and Chemla's (2017) Experiment 2. Adapted from "Color adjectives, standards, and thresholds: an experimental investigation" by N. Hansen and E. Chemla, 2017, *Linguistics and Philosophy, 40*, p. 252. Copyright 2017 by The Author(s). Reprinted with permission (http://bit.ly/2M0lpti).

The participants were asked to click on the tall alien in three different visual contexts: pictures 0/3 and 1/3, pictures 1/3 and 2/3, and pictures 2/3 and 3/3 were presented together. Interestingly, the participants selected only in the 2/3 vs. 3/3 condition the taller object. In the 0/3 vs. 1/3 and the 1/3 vs. 2/3 condition the participants responded that neither of the aliens is tall. Therefore, Hansen and Chemla (2017, p. 20) propose a standard "which is composed of three elements: a *lower threshold*, an *upper threshold*, and a *standard*". This type of standard is more abstract than the notion of standard described in Chapter 2 but this approach resembles the definition of the standard as a range of degrees (see Solt, 2011). In the case of positive gradable adjectives such as *tall*, objects have to cross the lower threshold on the corresponding scale before they are interpreted as having a context-dependent standard. As an example, Hansen and Chemla (2017) mention a situation with two tiny soldiers, with one noticeably taller than the other. The request *Please give me the tall soldier* may be reasonably objected because none of the soldiers is tall enough to count as a tall soldier. As a consequence, according to Hansen and Chemla (2017), people fail to adjust the standard at the lower end of the scale. In order to use a relative standard, that can be adjusted, the objects must be associated with a degree that is between the upper

[9] The labels mean that the 3/3-alien has the property to a maximal extent. Since this does not exist for *tall*, this was the alien which filled the box completely. The 1/3- and 2/3-alien was 1/3 or 2/3 height of the box. The 0/3-alien could not have a zero degree of height, hence it was depicted as extremely short

and lower threshold (Hansen & Chemla, 2017). However, note that the objects used in Hansen and Chemla (2017) are not perfectly comparable with the ones used in Barner and Snedeker (2008). It is likely that participants, especially adults, have some mental image of aliens (and their height). This might have influenced the participants' reactions as well.

Nevertheless, Hansen and Chemla's explanation can be applied to the results of the present study. The interaction of the adjective's polarity and the polarity of the visual expansion of the scale can be explained as follows: when participants hear the test prompt *Please give me the big toys* in a lower expansion context, the additional smaller objects are not taken into account because they are below the lower threshold for *big*. As a consequence, the average size of the comparison class does not change and the standard need not to be adjusted. The same holds for *small* in the opposite direction: when participants hear the test prompt *Please give me the small toys* in an upper expansion context, the additional bigger objects are not taken into account because they exceed the upper threshold for *small*.

The effect of the comparison class on the interpretation of gradable adjectives: non-target-like interpretations

Part 2 and 3 of the experiment also investigated the question whether interpretation patterns with respect to the relevance of the comparison class change with age (C2b.). The results showed that 3-year-old children, in contrast to 4- and 5-year-old children, adjusted the standard equally often for RGs and AGs. This finding is in line with hypothesis H1 because there is a stage prior to the target-like interpretation in which children do not distinguish RGs and AGs. In Section 5.1, I outlined that this pattern can result from three different interpretations of gradable adjectives. First, RGs may be the default interpretation for all gradable adjectives. Second, AGs may be the default interpretation for all gradable adjectives. Third, neither RGs nor AGs are interpreted target-like. I argue for the third option because of the following reasons. The first option would follow the approach by Toledo and Sassoon (2011) explained in Chapter 2. They argue that the value of the comparison class is relevant for all gradable adjectives. According to their approach, the difference between RGs and AGs results from their different scale structures and from distinct structures of the comparison classes. The comparison class for AGs includes different possible temporal stages of the same individual. In contrast, for the interpretation of RGs one individual is compared to other individuals, hence the comparison class for RGs includes only other members of the same category. The visual arrays presented in the experiment matches only the characteristics of the comparison class for RGs, i.e., it shows different members of the same category marked by different colors of the objects. To match the characteristics of the comparison class for AGs, the same objects in different temporal stages must have been shown. It is possible that older children and adults recognized that AGs refer to the latter kind of comparison class. Because this kind of comparison class was not presented in the experiment, older children and adults did not adjust the standard of comparison in AG-trials. It may be that the nature of the comparison class matters

less for 3-year-old children. As a consequence, they adjusted the standard for AGs even if the visual array matched only the comparison class for RGs. Because visual contexts showing comparison classes appropriate for AGs were not presented in the experiment, this conclusion remains purely speculative.

One could argue for the second option because the standard adjustments for RGs and AGs looked different. Therefore, the standard adjustments for AGs may not result from an interpretation of AGs dependent on a comparison class. Rather, it may be that children start with a comparison class-independent interpretation of RGs and AGs. Then, the standard adjustments for RGs can be explained by the developing knowledge about the relevance of the comparison class for RGs and the developing sensitivity to different kinds of gradable adjectives. However, the number of standard adjustments differed significantly between 3-year-olds and adults for both RGs and AGs. This finding implies that as of age 3, neither RGs nor AGs are interpreted adult-like with respect to the relevance of the comparison class. Therefore, I argue for the third option and claim that children around age 3 interpret gradable adjectives similar to semantically less complex non-gradable adjectives. The semantic representation of gradable adjectives at this stage is explained more closely in Chapter 6.

At first sight, the present findings seem to conflict with previous findings by Syrett (2007) and Syrett et al. (2006). The authors found that RGs, but not AGs, are interpreted dependent on the context. On the basis of this finding, they conclude that as of age 3 children distinguish RGs and AGs. The different findings may result from the different kinds of comparison classes investigated and the different tasks used in the experiments. In Syrett (2007) and Syrett et al. (2006) the comparison class was implicit. The test prompts were of the form *Please give me the big one*, hence the adjective did not modify a noun and the visual context provided the comparison class. In the present study, the comparison class was explicit, i.e., children had to use the information provided by the noun to detect changes in the comparison class because the visual context did not change. It is possible that visual cues are more prominent than linguistic cues. As a result, children may have performed better in the experiments by Syrett (2007) and Syrett et al. (2006). In addition, in these experiments children had to select the object matching the test prompt out of only two objects. In the present study, the children had to select the objects matching the test prompt out of twelve objects. It is likely that the present task was more demanding than the task used in Syrett (2007) and Syrett et al. (2006). Therefore, the children may have performed less adult-like than in the tasks used in previous studies.

Although changes in the noun did not affect the 3-year-olds' interpretation of gradable adjective in a target-like way, I assume that children took the noun into account when they performed the task. If not, different types of anwers would have been found. For instance, participants could have shown a general strategy to always select the same number of objects independent of the test prompt. Moreover, if Ninio's (2004) explanation holds for German, namely that children have problems integrating the adjective and noun meaning, more instances of the following answer should have been found: in Part 3, the visual array

contained objects from two basic-level categories, e.g., water balloons and bigger soccer balls. In the test prompt the adjective modified a basic-level noun, e.g., *Please give me the big water balloons.* If the noun is not processed, children should have taken all objects into account, i.e., water balloons and soccer balls. As a result, they should have selected also the soccer balls because they are big, too. Because this kind of response did not occur, I assume that German-speaking children are able to integrate adjective and noun meaning. Ninio's (2004) findings may be attributed to the Hebrew word order, which is different from German because the adjective follows the noun.

The effect of the comparison class on the interpretation of gradable adjectives: alternative interpretations

As shown in Section 5.5.2.3, participants did not respond to RG-trials in an all-or-none pattern because four different response patterns occurred as alternative to the expected standard adjustments. First, sometimes participants adjusted the standard of comparison but not in the expected direction given the distribution of objects in the visual array. This response type suggests that the participants took the comparison class into account, but they did not use the statistical information provided by the comparison class appropriately to calculate the standard of comparison. This means they did not take differences in the mean size of objects into account depending on whether the scale is expanded at the upper or at the lower end. Hence, they failed to detect wether the standard must shift towards the upper or the lower end of the scale. Findings by Barner and Snedeker (2008) indicate that 4-year-olds are sensitive to changes in the mean size of the object array. However, distributional properties other than differences in mean size, e.g., linear increasing object sizes *versus* object sizes divided into clusters, do not affect 8-year-old children's calculation of the standard (Booij & Sassoon, 2014). Therefore, difficulty with conceiving the statistical properties of the comparison class is not ruled out as a source of this unexpected standard adjustment pattern. This difficulty is probably more strongly connected to non-linguistic than to linguistic abilities. A relation to cognitive skills, processing capacities, or neurological maturation processes is likely. Siegel (1977) found that cognitive skills regarding the perception of number are not fully developed until age 4. Tribushinina (2013) notes the that interpretation of gradable adjectives may be related to executive functions, which develop between 3 and 6 years, and are associated with the prefrontal cortex. According to Waltz et al. (1999), the prefrontal cortex may be the locus of the system for relational reasoning in humans. One example for relational reasoning is to draw transitive inferences such as *If A is taller than B and C is taller than A, then C is taller than B.* According to Waltz et al. (1999), more reasoning is required to comprehend and integrate such multiple relations than to comprehend single relations such as *A is taller than B.* The establishment of multiple ordering relations between objects was also a necessary ability to succesfully perform the task of the present study. The prefrontal lobe however is the last cortical region to reach full structural development and increases slowly until the age of 8

(Kanemura, Aihara, Aoki, Araki, & Nakazawa, 2003). As a result immature cognitive skills and neurological processes may be a source of children's unexpected standard adjustments.

The second alternative response was the calculation of two different standards for each basic-level category presented in the visual context, when the test prompt contanined the superordinate-level noun *toys*. This response type also suggests that participants took the comparison class into account for the interpretation of RGs. The comparison class encoded by the superordinate-level noun *toys*, a second-order predicate, comprised two different basic-level categories. Thus, a test prompt containing the noun *toys* can be interpreted in two ways. Consider for example the test prompt *Please give me the big toys*. This request can be interpreted as *Please give me the objects which are big for toys* if the participants abstract away from the differences between the two basic-level categories and combine them to one superordinate-level category. In contrast, the test prompt can be interpreted as *Please give me the toys which are big for X (e.g., water balloons) and the toys which are big for Y (e.g., soccer balls)* if the participants maintain the two basic-level categories and understand them as two different kinds of toys. The computation of two separate standards for the respective subcategories of toys as in the latter example could have been triggered by a general preference for basic-level categories because this is the easiest level of abstraction for children and adults (e.g., Mervis & Rosch, 1981). It has been argued that basic-level categories are cognitively most efficient because they maximize within-category similarity relative to between-category similarity. In addition, at the basic-level one mental image can reflect the entire category (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Early studies in psychology and language acquisition have shown that participants recognize objects more rapidly as members of a basic-level category than as members of categories at other levels of categorization. In addition, participants use basic-level terms to name objects spontaneously (e.g., Rosch et al., 1976). Callanan and Markman (1982) found that children sometimes interpret *toy* as a group of objects that are related to each other in some way, but refuse to call single objects that belong to this group (e.g., a doll) *toy*, although they share the criteria for category membership. Sometimes this interpretation is maintained even by adults (Murphy & Wisniewski, 1989). Moreover, Gelman and O'Reilly (1988) pointed out that children have less difficulty with understanding superordinate-level nouns when they refer to natural kinds rather than to artifacts like toys because artifacts are related functionally rather than taxonomically. Regarding the present experiment, it is open why two different standards for the two basic-level categories occurred less often in the 3-year-old group compared to the 4-year-old, 5-year-old, and adult group. It is likely that calculating two standards, one for each subcategory of toys, is cognitively more demanding than calculating only one standard.

As the third alternative response participants did not adjust the standard of comparison for RGs. This response type is possibly related to the within-subjects design of the experiment. The present design has the advantage that it allowed to investigate whether changes in the noun influenced the participants' calculation of the standard of comparison for the identical visual array of objects. The disadvantage of a within-subjects design is

that responses to one trial may be carried over to another trial. This is a likely outcome if the visual context is the same for both trials. This influence is spelled out in a comment by one adult participant, who said that she has to remember the size to which she defined the objects as small ("Ich muss mir immer merken, bis zu welcher Größe ich das als klein definiert hab"). This comment illustrates that it seems that the participant tried to imagine what the task was, which in turn led to the unexpected response of not adjusting the standard for RGs.

Another property of the design could also be responsible for not adjusting the standard: it is possible that the differences in size of the depicted objects were too small to induce a standard shift. As a consequence, from the participants' view it may have been unnecessary to calculate different standards for the basic-level and the superordinate-level category. In other words, that the participants are willing to judge for instance the big water balloons also as big toys. Sera and Smith (1987) report that children's standard shifts for gradable adjectives such as *big* increase the bigger the difference between the objects are. Booij and Sassoon (2014) claim that people make distinctions, e.g., between big and small, where and when they think it is worth making this distinction. Moreover, participants prefer distinctions that are easy to make. This behavior results in the Sorites Paradox described in (2.53). To explain this unwillingeness to make distinctions between entities that differ only slightly, Booij and Sassoon (2014, p. 2) introduce the framework of semi-orders. Semi-orders are defined as "an ordering relation between items that is less strong than a weak order, which is the sort of order that obtains when the whole domain is thought of as partitioned into degrees". An example for a weak order is the *taller than*-relation. An example of a semi-order is the *much taller than*-relation: for instance, John is not much taller than Susan, Susan is not much taller than Mary, but John can be much taller than Mary (Booij & Sassoon, 2014, p. 2). According to this approach, there is an 'indistinguishability threshold' below which no "real" inequality exists. Choosing a semi-order at the arrangement of objects can be used as a first and simple way of classifying them. If this does not yield a standard of comparison, a more fine-grained, degree-based ordering becomes necessary. Regarding the present results, one could suppose that the participants made a distinction between small and big objects for the first distribution they saw and for the first noun they heard, which was *toys*. When they saw the same distribution of objects for the second time, it may be that they considered the same objects big or small, although the adjective modified a basic-level noun. Because the differences in size between the depicted objects were not large enough, people did not assess the objects as inequal. Furthermore, the basic-level category is part of the superordinate-level category: water balloons for instance are toys, hence there was no reason for the participants to shift the standard of comparison.

The fourth alternative respose were unanalyzable answers. The number of unanalyzable answers decreased with age when the interpretation of gradable adjectives is more mature. This response type can be explained by children's initial non-target-like semantic representation of gradable adjectives, which is explained more closely in Chapter 6.

In summary, Part 2 and 3 of the experiment indicate that children differentiate between relative and absolute gradable adjectives regarding the relevance of the comparison class from age 4. This sensitivity to the relevance of the comparison class develops between the age of 3 and 4 and showed up in different response patterns. For the 3-year-old children, the findings indicate no difference between RGs and AGs. This result points to a non-adult-like interpretation of both RGs and AGs. In the last section of this chapter, I summarize and relate the findings of Part 1, 2 and 3 of the experiment.

5.6 Developmental steps in the comprehension of gradable adjectives

This section summarizes the developmental steps for the acquisition of gradable adjectives between age 3 and 5. The focus of the present study was the interpretation of relative and absolute gradable adjectives regarding the standard of comparison (C1) and the relevance of the comparison class (C2). Because RGs and AGs do not differ in semantic complexity, H1 predicts that they are interpreted target-like at the same age, i.e., that children interpret them differently. The target-like interpretation is preceded by the same interpretation for both RGs and AGs. In the following, I describe the developmental steps in the comprehension of gradable adjectives as observed in the present study. In a first step, knowledge about the distinction between RGs and AGs emerges: with regard to the position of the standard on the scale, RGs and AGs are distinguished target-like. With respect to the nature of the standard and the relevance of the comparison class, the distinction is not yet target-like. Regarding the standard of comparison children distinguish between RGs and AGs such that they locate them at different positions on the corresponding scale. For adjectives that have (partially) closed scales, i.e., absolute gradable adjectives, the standard of comparison is either the minimum or maximum endpoint of the scale. For adjectives that have open scales, i.e., relative gradable adjectives, the standard of comparison is located around the center of the scale. For RGs and AGs, the standard of comparison constitutes a single point, or degree, on the scale, which is the same for both the negative and the positive antonym. Regarding the relevance of the comparison class, RGs and AGs are interpreted alike; neither of them is interpreted adult-like. At this stage, many unanalyzable answers occurred in the data.

In a second step, children have acquired that the comparison class is relevant for the interpretation of RGs, but not for the interpretation of AGs. A prominent response type was to adjust the standard for RGs, but in the unexpected direction. This response type indicates that children may not be able to conceive the statistical properties of the visual array properly. In addition, children become aware that the standard of comparison for RGs, but not for AGs, is different for negative and positive antonyms. As a consequence, they calculate two different standards, which results in answers in which the two standards were reversed.

In a third step, children approach the adult-like interpretation of gradable adjectives. The standard of comparison for RGs is defined as a range of degrees indicated by a gap between positive and negative antonym pairs. RGs are interpreted relative to the comparison class especially in particular combinations of adjective and visual context: positive adjectives were more likely to trigger a standard shift when the scale was expanded at the positive pole, whereas negative adjectives were more likely to trigger a standard shift when the scale was expanded at the negative pole.

This chapter showed that children at age 3 distinguish between relative and absolute gradable adjectives when defining the standard of comparison. This finding adds to previous research evidence from a different experimental task and from another language. In addition, absolute and relative gradable adjectives of different polarity were tested. The present study extends previous work by investigating children's interpretation of explicit comparison classes for relative and absolute gradable adjectives. It further broadens previous findings by investigating different distributions of objects. A greater number of items and participants is necessary to support the conclusions presented in this section.

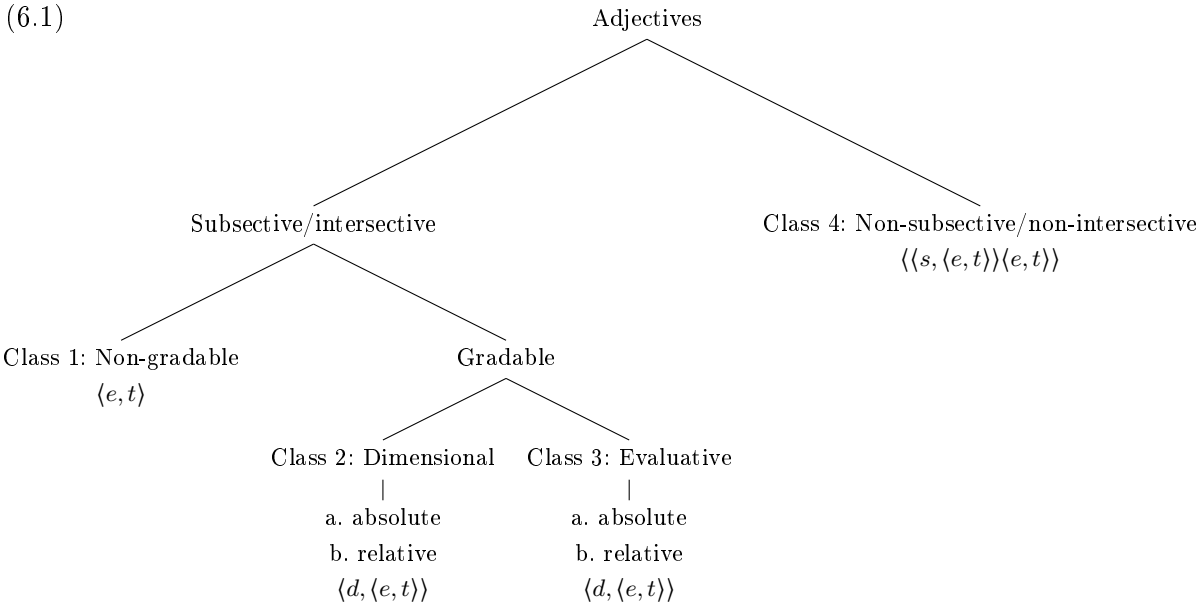
The finding that the target-like distinction between RGs and AGs with regard to the standard of comparison is acquired earlier than the distinction between RGs and AGs with regard to the relevance of the comparison class calls for an explanation. In the next chapter, I discuss how the developmental steps identified in this study relate to each other and to the findings of the spontaneous speech study reported in Chapter 4. I further assess how the developmental sequence relates to the notion of semantic complexity and what children have to acquire to move on to the following stage.

Chapter 6

General discussion

This thesis addresses the question of how children acquire the semantics of adjectives. In particular, this question concerns the common question in language acquisition research in which order children acquire lexical categories, in this case adjectives, and which factors influence this order of acquisition. In Chapter 2, I proposed a classification of adjectives based on the entailment and gradability properties proposed in the semantic literature (see (6.1)) and concluded that the adjective classes differ in semantic complexity. I defined semantic complexity as a combination of the adjective's length of description, the complexity of its semantic type, and additional lexical information that contributes to the adjective's meaning. The differences in semantic complexity are expressed by the *Semantic Complexity Hierarchy* in which Class 1 is least complex and Class 4 is most complex. Regarding the acquisition of the semantics of adjectives, I hypothesized that the acquisition order of adjectives is determined by their semantic complexity.

(6.1)



A spontaneous speech data analysis (see Chapter 4) and a comprehension experiment (see Chapter 5) were performed to examine whether the *Semantic Complexity Hierarchy* is reflected in the acquisition order of adjectives. In other words, is the acquisition order

of adjectives determined by their semantic complexity? As shown in Chapter 3, previous studies mention two alternative factors possibly affecting the order of acquisition: notional properties of adjectives, which relate to the concepts adjectives express, and adjective frequency in the input. These two factors were also addressed in this thesis with regard to the order of acquisition in spontaneous speech. By combining the empirical findings of this thesis with the findings from semantic theory and previous acquisition studies I propose a developmental path for the acquisition of adjectives.

This chapter is organized as follows: in Section 6.1, I summarize the main findings of the spontaneous speech data analysis and the comprehension experiment in light of the hypotheses and research questions presented in Chapter 2 and 3. A developmental path for the acquisition of adjectives resulting from semantic complexity is proposed in Section 6.2. I suggest that children start with an interpretation of all adjectives as properties of individuals. The first distinction between adjective classes is acquired when gradable adjectives, in contrast to non-gradable adjectives, are interpreted as relations between individuals and degrees. Finally, children differentiate between intersective and non-intersective adjectives as soon as they are able to interpret non-intersective adjectives as functions from properties to properties. In Section 6.3, I discuss further implications of this acquisition path for other classification systems of adjectives, for the acquisition of adjectives across languages, and for the acquisition of other modification structures.

6.1 What determines the acquisition order of adjectives?

Complexity has been claimed to predict the order of acquisition in the area of morphosyntax: structures that require more complex derivations should not be acquired before structures that require less complex derivations (Brown & Hanlon, 1970; Frank, 1998; Hawkins, 1987; Moscati & Rizzi, 2014; Snyder, 2007). To date, semantic complexity has received little attention in acquisition research, in particular with respect to the production and comprehension of adjectives. Previous research on adjectives in spontaneous speech identified factors such as the abstractness of the concepts underlying adjectives or the adjectives' frequency in the input as predictors for the acquisition order. Studies that considered semantic properties as a predictor for the acquisition order in spontaneous speech were restricted to particular adjective classes such as gradable adjectives. In this thesis, a comprehension and a production study were performed to test the hypothesis (H1) that semantic complexity determines the order of acquisition of adjectives.

If semantic complexity determines the acquisition order of adjectives as stated in H1, it is predicted that adjectives that are more complex in the *Semantic Complexity Hierarchy* are not acquired earlier in spontaneous speech than adjectives that are less complex. Adjectives of the same complexity should be acquired simultaneously. For the longitudinal spontaneous speech study of this thesis a dense data corpus of one German-speaking child between 2;00 and 2;11 years was analyzed. In total, 202 of all adjectives the child produced were assigned to one of the classes in (6.1). For these adjectives, research ques-

tion (P1) asked whether the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition of the adjective classes. As predicted by H1, adjective classes of different complexity differed in their mean age of acqusition – more complex adjective classes were not acquired before less complex adjective classes – and adjectives of the same complexity did not differ in their mean age of acquisition. This order was observed especially for adjectives produced in combination with nominals, i.e., in attributive and predicative structures. The most complex adjective class was not attested until age 2;11. Hence, the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition of the adjective classes. Research question (P2) asked whether the *Semantic Complexity Hierarchy* is reflected in the growth patterns of the adjective classes. The definition of growth is taken from the notion of vocabulary spurt or vocabulary explosion that has been observed in lexical acquisition around age 2 (e.g., E. V. Clark, 1993; Menyuk et al., 1995). This phenomenon describes that children acquire a large amount of new words in a short period of time. It is conceivable that this phenomenon may also appear for specific word classes, in this case adjectives. With regard to complexity, H1 predicts that for adjectives of the same complexity, this "explosion" happens at the same age, whereas for adjectives of different complexity it should happen at different ages and should be sequenced according to the classes' complexity. The adjective vocabulary explosion was measured as the maximal number of new types. As predicted by H1, more complex adjective classes "exploded" after less complex adjective classes, and adjectives of the same complexity "exploded" at the same age. For the class of intersective/subsective relative gradable dimensional adjectives (Class 2b 'BIG') two peaks were found: this class "exploded" together with the equally complex class of intersective/subsective absolute gradable dimensional adjectives (Class 2a 'CLEAN') and together with the less complex class of intersective/subsective non-gradable adjectives (Class 1 'BLUE').

Taken together, the findings for adjectives in spontaneous speech indicate that first, on average at 2;04 years, subsective/intersective non-gradable adjectives are acquired as (attributive) modifiers of nouns followed by subsective/intersective dimensional gradable adjectives at 2;06 years and subsective/intersective evaluative adjectives at 2;08 years. Non-subsective/non-intersective adjectives are the last class to be acquired; they were not produced during the time period investigated. The adjective vocabulary explosion happened in the same sequence around three months before the mean age of acqusition for each class. These sequences support the hypothesis (H1) that the order of acquisition of adjectives is determined by their semantic complexity. What about the input as another factor possibly affecting the acquisition of the semantics of adjectives and the acquisition order?

The role of the input for the acquisition order of adjectives is not resolved. Although there is no doubt that the language learner must hear an adjective before an entry for that word is created in the mental lexicon, it is far from clear whether the input actually determines the order of acquisition and provides the evidence the child needs to acquire the subtle differences between the adjective classes. Hypothesis H2 states that the acquisition

order of adjectives is not completely determined by their frequency in the input. Therefore, H2 is not a real alternative hypothesis to H1, but rather an addition to H1; it acknowledges the input as a trigger for establishing the adjective classes in the *Semantic Complexity Hierarchy*. If the order of acquisition of adjectives would be completely determined by their frequency in the input, more frequent adjectives in the input should be acquired earlier than less frequent adjectives, or at the same age. Adjectives that are equally frequent should be acquired simultaneously. Hence, if H2 is confirmed, we should not find a systematic correlation between the frequency of adjective use in the input and the age of acquisition. Concerning the role of the input, the questions were whether and how frequency of adjective use in child speech and in the input are correlated (P3), whether the frequency of adjective use in the input is reflected in the age of acquisition (P4), and whether late acquired adjectives are missing in the input before the age of acquisition (P5). The data showed a significant positive correlation between frequency of adjective use in the input and in child speech (P3), i.e., the more frequent an adjective was used in the input the more frequent it was used by the child. However, not all semantic classes were equally distributed in terms of types and tokens across the child and his caregivers. This is predicted by H2: it is not surprising that the adjective use in the speech of the child and his caregivers influence each other, but the child's adjective use is not completely dependent on his caregivers' adjective use and *vice versa*. Regarding (P4), the data showed a significant negative correlation between frequency of adjective use in the input and the age of acqusition, i.e., the more frequent an adjective was used in the input the earlier it was acquired by the child. This finding suggests that the order of acquisition is determined by the frequency in the input, apparently contradicting H2. However, based on five specific adjectives that were acquired late, at age 2;11, I illustrated that these adjectives are nevertheless used in the input from the beginning of the corpus study at age 2;00 (P5). Moreover, five additional adjectives with a similar input frequency were acquired early. These findings support H2 because adjectives were acquired late although they were present in the input a considerable time before the age of acquisition. In addition, similarly frequent adjectives were not acquired simultaneously.

The findings for research questions (P1) to (P5) of the spontaneous speech data analysis indicate that the acquisition order of adjectives is determined by their semantic complexity and is partially influenced by their frequency in the input. With regard to both hypotheses further research is necessary. More children should be investigated and the age range should be extended to younger and older ages to detect the very first adjectives produced and to identify the age at which the most complex class of non-subsective/non-intersective adjectives are acquired. The input should be analyzed more carefully with regard to the syntactic environments in which adjectives are produced to find out whether there is evidence for the child to infer the adjectives' meanings. In addition, to interpret the correlations between adjectives in the input and in child speech properly, it is necessary to analyze in detail whether the caregivers provide the adjectives to the child or *vice versa*.

Another crucial finding of the spontaneous speech data analysis concerns the relation

between the semantic classes investigated in this thesis and the notional classes investigated in previous studies. The adjectives produced by the child were assigned to one of the semantic classes as well as to the notional classes employed in previous studies, allowing to explore a correspondence between the two classification systems. I propose that notional and semantic classes are related in the following way: notional classes relate to the concepts adjectives express. Semantic classes refer to the compositional and lexical semantic properties underlying these concepts. Thus, notional properties are important for the constitution of the mental lexicon and semantic properties are relevant for the adjectives' function as modifiers, i.e., their combination with nouns. Therefore, the two classification systems should be evaluated as complementing one another. How the language learner can benefit from both notional and semantic properties in the acquisition of adjectives is shown in Section 6.3.

The role of semantic complexity was also addressed regarding the comprehension of adjectives. The comprehension experiment in this thesis focused on two specific adjective classes: relative (*big, small*) and absolute (*clean, dirty*) gradable dimensional adjectives (Class 2a and 2b). These two classes are relevant for the notion of semantic complexity in acquisition for two reasons. First, according to the definition of semantic complexity I proposed in Chapter 2, they are supposed to be more complex than non-gradable adjectives (Class 1 'BLUE') because they denote relations between individuals and degrees rather than just properties of individuals. Hence, they are a case in point for the question how children interpret these adjectives and how they figure out that adjectives are gradable. Second, relative and absolute gradable dimensional adjectives are supposed to be equally complex. However, although they have the same semantic type, they are interpreted differently because of lexical semantic properties such as their scale types. With regard to the acquisition of these two classes, this thesis investigates how children come to differentiate adjectives that have the same semantic type but differ in their interpretation. The differences in interpretation can be observed with regard to the standard of comparison and the relevance of the comparison class for determining the standard. To date, there are few studies comparing the interpretation of relative and absolute gradable adjectives. Previous studies indicate that by age 3, children distinguish between relative and absolute gradable adjectives with regard to the standard of comparison and the relevance of implicit comparison classes. Implicit comparison classes are not linguistically encoded by the modified noun but by visual context only. Thus, it is open whether children can use the presence of a noun to determine the comparison class. Furthermore, it is open for which gradable adjectives changes in the noun affect the interpretation. In addition, because mostly positive pole gradable adjectives (e.g., *big*) have been studied, it is open how negative pole gradable adjectives (e.g., *small*) are interpreted in relation to their positive counterparts.

The comprehension experiment in this thesis was inspired by the design of Barner and Snedeker (2008). They manipulated the comparison class by using different nouns, i.e., different explicit comparison classes, without changing the visual context. The original design

was modified in five crucial respects. First, the present experiment included both positive and negative relative gradable (*groß* 'big', *klein* 'small') and absolute gradable (*sauber* 'clean', *dreckig* 'dirty') adjectives rather than relative gradable adjectives only. Second, a within-subjects design instead of a between-subjects design was used to assess the direct influence of changes in the comparison class on the participants' interpretation. Third, the adjectives modified existing nouns instead of novel nouns. This third modification is related to the fourth modification that the age range of the participants was extended to younger and older children. Barner and Snedeker (2008) tested 4-year-old children, in the present comprehension experiment 3- to 5-year-old children were tested. Using existing nouns is relevant to examine whether knowledge about existing objects affects the interpretation of gradable adjectives. This possibility was mentioned by Barner and Snedeker (2008) with regard to the interpretation of children younger than age 4, who they did not test. Fifth, the objects in the visual context were presented without any ordering with the consequence that the participants had to establish their own ordering before calculating the standard of comparison. This way, the participants were prevented to infer their judgements from an already given order.

For adjectives that are equally complex, hypothesis H1 predicts that they should receive a target-like interpretation at the same age. An important question is how children interpret relative and absolute gradable adjective before they interpret them target-like. It may be that children start with the same interpretation for all gradable adjectives. 'Same' in this respect can mean that either relative or absolute gradable adjectives are selected as the default interpretation. However, 'same' can also mean that neither relative nor absolute gradable adjectives are interpreted target-like initially.

To test hypothesis H1, typically-developing monolingual German-speaking children between 3 and 5 years were tested. Research question (C1a.) asked: Do children differentiate between absolute and relative gradable adjectives when defining the standard of comparison? The findings indicate that 3- to 5-year-old children do not differ from the adult controls. Both groups selected a degree around the center of the scale as the standard for relative gradable adjectives (*big, small*), a non-zero degree for minimum absolute gradable adjectives (*dirty*), and the maximal degree for maximum absolute gradable adjectives (*clean*). To find out whether there is a developmental stage at which children do not differentiate between relative and absolute gradable adjectives target-like, it was also investigated whether the standard of comparison changes with age (C1b.). The findings indicate that from age 3, children differentiate between absolute and relative gradable adjectives when defining the standard. Thus, this meaning component is acquired at the same age for both relative and absolute gradable adjectives as predicted by H1. It is open whether there is a stage at which this difference is not settled and which response patterns children would show at this stage. Although the position of the standard is calculated target-like already at age 3, the findings indicate developmental changes with respect to the nature of the standard for relative gradable adjectives. Identifying these changes was only possible because positive and negative adjectives were investigated. Initially, the standard for relative

gradable adjectives can be characterized as one degree that is shared by both antonyms. In a next step, children establish a gap between positive and negative adjectives resulting in two different standards for antonym pairs. Finally, the standard can be analyzed as a range of degrees (see Solt, 2011). The standard for absolute gradable adjectives is defined target-like as a single degree from the outset. Hence, as predicted by H1 there is a stage at which absolute and relative gradable adjectives are interpreted similarly regarding the nature of the standard.

Research question (C2a.) asked: Which gradable adjectives do children interpret relative to a comparison class? The results suggest that both the group of 3- to 5-year-olds and the adult control group interpret relative gradable, but not absolute gradable adjectives, relative to a comparison class. This means that changing the noun results in an adjustment of the standard for relative gradable adjectives. However, this interpretation pattern changed with age (C2b.). The group of 3-year-old children differed from the adults with respect to the interpretation of both absolute and relative gradable adjectives. As of age 4, children did not differ in their interpretation of absolute and relative gradable adjectives from adults. These findings are in line with H1 because absolute and relative gradable adjectives are interpreted target-like at the same age, at age 4, and receive the same non-target-like interpretation prior to that age.

An analysis of children's responses that did not meet the expected response revealed more unalyzable answers in the group of 3-year-olds than in the other age groups. In the group of 4-year-olds, standard adjustments that were not expected by the distribution of objects in the visual array were found more often than in 3- and 5-year-olds and in the adult group. Summarizing the findings of the comprehension experiment yields the three developmental steps in Table 6.1.

Table 6.1: Developmental steps in the comprehension of relative and absolute gradable adjectives.

Step	Age	Standard of comparison	Relevance of the comparison class
1	around 3	Standard is located at different positions on the scale (endpoint vs. midpoint) Same standard degree for antonyms of RGs and AGs	Non-target-like interpretation of RGs and AGs
2	around 4	Two different standards for antonyms of RGs	Distinction between RGs and AGs Adjustments of standard for RGs not as expected by the distribution of objects
3	around 5	Standard as a range of degrees for RGs	Adjustments of standard for RGs sensitive to distribution of objects

Note. RG = relative gradable adjective, AG = absolute gradable adjective.

In summary, the findings of both studies indicate that semantic complexity affects the acquisition order of adjectives in production and comprehension. Further research is necessary to investigate the role of other factors such as the input for the acquisition of

adjectives. Investigating children at different ages revealed that the target-like interpretation of adjectives follows their production – an observation that supports previous findings on adjectives (Hohaus et al., 2014; Tribushinina & Gillis, 2012) and on other modification structures such as relative clauses (Friedmann & Novogrodsky, 2004), prepositional phrases (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000), and focus particles (Müller, 2012).

In the next section, I combine the findings of the two studies of this thesis with findings from previous research and propose a developmental path for the acquisition of adjectives.

6.2 Stages in the acquisition of adjectives

In Section 6.2.1, I start with an overview of aspects mastered and not mastered in the time course of the acquisition of adjectives. Focusing on the comprehension of adjectives at acquisition stages 3 to 5, I propose what the semantic representation of adjectives at each stage may be. In addition, I illustrate how the semantic representation in question can account for the aspects mastered and not mastered. In Section 6.2.2, I relate the acquisition path based on findings from comprehension to findings from production. In Section 6.2.3, I discuss how children can reach the next stage given their inherent linguistic knowledge and the positive evidence from their linguistic environment.

6.2.1 The semantic representation of adjectives in the time course of acquisition

Previous studies on the acquisition of adjectives cover roughly the age range between the 50 word phase and 5 years. These studies indicate that the language learner masters various aspects of the acquisition task during this time. The language learner recognizes novel words as adjectives and assigns the correct meaning, i.e., denoting properties of objects. In addition, the child acquires the meaning of specific lexical items. Another task is to combine adjectives with other elements such as nouns. In order to interpret these combinations correctly, the child must integrate the meaning of adjective and noun. A further task is to find out that the noun can be of different relevance for the interpretation of adjective-noun combinations. The main findings of previous acquisition studies (see Chapter 3) are summarized in Tables 6.2 to 6.6 (in italics) and supplemented by the findings of this thesis (in bold). If a finding is written in italics and bold, it indicates that the present findings replicate previous findings. Recall that the indication of age in developmental sequences is understood as an approximate value because individual differences between children are possible. Crucially, the order of the acquisition stages is not expected to vary between children. If cells are empty then, to my knowledge, no empirical evidence exists.

Stage 1:

In Stage 1, children build their mental lexicon. That means they have to isolate word forms,

identify the possible meaning and link both together (E. V. Clark, 1993). The literature addressing the age of Stage 1, 1;06-2;00 years, mostly concerns children's expectations about the meaning of novel words. The question is whether children know that adjectives refer to object properties rather than to object categories like nouns typically do. This question results from the general issue of how children are able to identify word forms and to assign meaning to these forms. It has been proposed that children use the correlation between conceptual categories such as OBJECTS, ACTIONS, PROPERTIES and grammatical categories such as NOUNS, VERBS, ADJECTIVES ('semantic bootstrapping', Pinker (1984)). By age 1, children have a substantial repertoire of conceptual categories (E. V. Clark, 1993). However, the link between conceptual and grammatical categories is only indirect and moreover, grammatical categories are defined by their usage in the sentence. Hence, the syntactic context of a novel word can also be a cue for the word class and the corresponding meaning ('syntactic bootstrapping', Gleitman (1990)). For adjectives in German, the syntactic context can be either the position between the determiner and the noun and the inflectional morphology or the presence of a copula.

As can be seen in Table 6.2, although adjectives are among children's first 50 words, children map novel words in syntactic adjective contexts to both object properties and object categories. Regarding the production of first adjectives, it is difficult to determine whether they are used referentially and whether they express the grammatical categrory ADJECTIVE in the adult-language, because a syntactic context is lacking. According to E. V. Clark (1993), a word like *hot* for example can refer to the property of having a high temperature, but can be used as a word for "oven" or "Be careful! Don't touch it!".

Table 6.2: Overview over the aspects mastered and not mastered in the acquisition of adjectives at Stage 1.

		Comprehension		Production	
Stage	Age	Mastered	Not mastered	Mastered	Not mastered
1	1;06-2;00		*Interpretation of novel words occuring in syntactic adjective frames as label for both object category & object property*	***First adjectives as single-word utterances***	

Note. italic = Findings from previous studies, bold = Findings from studies of the present thesis.

Regarding the form-meaning mapping, children seem to have problems to identify novel words as object properties. The non-adult-like expectation regarding novel adjectives can have linguistic and non-linguistic reasons. The first linguistic reason for children's non-adult-like expectation regarding novel adjectives could be related to the semantic denotation. As illustrated in Chapter 2 and repeated in (6.2), just like nouns and verbs, adjectives denote properties of individuals (type $\langle e, t \rangle$).

(6.2) a. Dumbo fliegt. λx.(**fliegen**(x))
'Dumbo flies.'

b. Dumbo ist ein Elefant. λx.(**Elefant**(x))
'Dumbo is an elephant.'

c. Dumbo ist männlich. λx.(**männlich**(x))
'Dumbo is male.'

It is open how children master the distinction between categories that have the same extension, namely sets of individuals with the respective property (flying, being an elephant, being male). It may be that the distinction between word classes is triggered by conceptual categories. Therefore, the second reason for children's non-adult-like form-meaning mapping may be that the relation between conceptual and grammatical categories is not explicit enough for the child. As a consequence, it may take some time to establish the link between the word form and the word meaning.

Very young children also fail to map novel words to object properties although they are presented in a syntactic adjective context (e.g., *this is a blikish horse* or *this horse is blikish*). Hence, the third linguistic reason may be that children have not yet enough words or the necessary words in their lexicon to infer the meaning of the novel word from the context. It may also be that the child is not familiar enough with the distribution of determiners or inflectional morphology. However, this possibility is unlikely because already very young children (14 months) map novel words presented in syntactic noun contexts (e.g., *this is a blik*) exclusively to the object category (Waxman & Booth, 2001).

Fourth, if children's form-meaning mapping is guided by lexical constraints (Markman, 1990), it may be that some constraints are more prominent than others to the child. One candidate constraint is the *whole-object assumption*. According to this constraint, the child should assume that a new word labels "the whole object and not its parts, substance or other properties" (Markman, 1990, p. 59). This constraint would explain why children are able to recognize nouns as labels for an object category, but not why children map novel adjectives to both object categories and object properties. It is possible that the conflict with other constraints such as the *mutual exclusivity assumption*, i.e., each object has one and only one label, enables children to acquire labels for object properties (Markman, 1990): if the child knows that *horse* is the label for the whole object, hence the category, she can infer that *blikish* must refer to something else, e.g., a property. However, it is also possible that the child relates *blikish* to the horse's tail rather than to a property to avoid a conflict of level.

As a non-linguistic reason, Smith, Gasser, and Sandhofer (1997) propose that children have difficulty to attend selectively to object properties. However, it is unclear whether and how the development of selective attention and the development of adjectives depend on each other.

Stage 2:

At Stage 2, two processes take place. The first process is related to the acquisition of the adjectives' lexical meaning. As can be seen in Table 6.3, in contrast to Stage 1, children acquire the ability to map adjectives to object properties. It may be that as of age 2, children are able to use the syntactic and semantic information of known words and the syntactic context to infer that adjectives refer to object properties. This process is easiest if the modified noun is a specific lexical NP (Mintz, 2005; Mintz & Gleitman, 2002). At the same time, children start producing adjectives in attributive and predicative structures, which points to knowledge about the adjective's behavior in combination with other lexical elements.

Table 6.3: Overview over the aspects mastered and not mastered in the acquisition of adjectives at Stage 2.

		Comprehension		Production	
Stage	Age	Mastered	Not mastered	Mastered	Not mastered
2	2;00-2;06	• *Interpretation of novel words modifying a specific lexical NP as object property* • *Integration of adjective and noun meaning* • *Ability to label objects as both 'big' & 'little' depending on the context if differences to other objects are large enough* • *Ability to switch from normative to perceptual standards*	• *Incremental use of adjective to identify referent* • *Standard for relative gradable adjectives at the extreme of the scale rather than at the midpoint*	• *Most notional classes* • **Most semantic classes** • *Distinction between gradable and non-gradable adjectives wrt degree modifiers* • ***Adjectives in attributive and predicative position***	• *Notional classes: temporal, internal state, age, behavior, modal, ordinals* • **Semantic classes: Non-subsective/non-intersective** • *Comparative forms of gradable adjectives*

Note. italic = Findings from previous studies, bold = Findings from studies of the present thesis.

In addition, children extend their mental lexicon with different subcategories of adjectives, i.e., different notional and semantic classes. The development of different subcategories of the adjective class seems to occur during the time of the vocabulary spurt in the development of the lexicon. It was found in the spontaneous speech data analysis in this thesis and in previous studies (Blackwell, 2005; Tribushinina et al., 2014) that the subclasses emerge in a specific order. This sequenced acquisition is similar to other word classes. For instance, in the class of nouns basic-level terms are acquired before super- and subordinate-level terms (Waxman, Shipley, & Shepperson, 1991). For verbs, it was found that telic verbs are acquired before atelic ones (Schulz, 2018).

Among the subcategories of adjectives are gradable adjectives. First relational interpretations of gradable adjectives, in contrast to non-gradable adjectives, are attested in comprehension. However, instances in which children do not interpret gradable adjectives

as relational are also found. I come back to the acquisition of gradable adjectives at the next acquisition stage. The distinction between gradable and non-gradable adjectives is visible in production because children modify gradable adjectives, but not non-gradable adjectives, with degree adverbs such as *very*. Nevertheless, comparative forms of gradable adjectives are not produced at Stage 2 indicating that children just start to acquire that adjectives can be gradable.

The second process at Stage 2 concerns the compositional semantics of adjectives. That is, that adjectives modify nouns and serve to restrict the set of possible referents. In this regard, the child's task is to integrate the adjective and the noun meaning, a process different from the process to map adjectives to object properties (see Fernald et al., 2010). At 2;06, children can use adjectives to identify the correct referent but they do not benefit from the presence of an adjective to find a referent more quickly. According to Fernald et al. (2010), it is not clear why this is the case. The authors propose processing efficiency, memory capacities, and skills in semantic integration as contributing factors.

Stage 3:

When children enter Stage 3, they know that adjectives denote properties of objects and they are able to integrate the adjective and the noun meaning. Because these conclusions are based on studies that used novel adjectives or color adjectives, the conservative assumption is that children's semantic representation of adjectives is the following: adjectives denote non-gradable properties of individuals (type $\langle e, t\rangle$) and combine with the noun extension via set intersection. At this acquisition stage, children also have adjectives from different classes in their productive vocabulary and we see the first relational interpretations of gradable adjectives. How do children acquire that adjectives are gradable, i.e., that they denote relations between individuals and degrees and that they have to be interpreted with respect to a standard of comparison? And, how do children come to differentiate gradable adjectives that have the same semantic type but differ in meaning? The second question may also be related to the acquisition of different scale types. It has been claimed in semantic theory that the scale type is related to the position of the standard of comparison and its dependence on a comparison class.

Because production data allow only indirect inferences about children's semantic representation of gradable adjectives, e.g., via the use of degree modifiers or comparison constructions (see Hohaus et al., 2014; Tribushinina & Gillis, 2012), I focus on the findings from comprehension experiments in Table 6.4 to 6.6 and come back to the findings from production studies in Section 6.2.2. On the basis of the findings for comprehension, I propose that children around age 3 interpret gradable adjectives as properties of individuals without reference to degrees.

It has been claimed in previous studies (H. H. Clark, 1970; Panzeri & Foppolo, 2012; Tribushinina, 2013) that initially, children interpret adjectives like *tall* as "having extent", i.e., they do not interpret these adjectives as relational or gradable. As a consequence, they either apply *tall* only to the tallest object in a series because this is the most typical

Table 6.4: Overview over the aspects mastered and not mastered in the acquisition of adjectives at Stage 3.

		Comprehension		Production	
Stage	Age	Mastered	Not mastered	Mastered	Not mastered
3	3;00	• *Ability to extend novel adjectives to objects from the same basic-level category (and across categories)* • *Interpretation of relative gradable adjectives that refer to more specific dimensions* • ***Distinction between relative and absolute gradable adjectives wrt standard** and perceptual comparison class* • **Standard for absolute gradable adjectives: antonyms share one standard degree**	• *Nominal interpretation of relative gradable adjectives instead of relative interpretation* • *Ability to label objects as both 'big' & 'little' if the visual context does not change* • **Standard for relative gradable adjectives as range of degrees: antonyms share one standard degree** • **Distinction between relative and absolute gradable adjectives wrt explicit comparison classes** • **Consistent ordering of objects** • *Interpretation of non-gradable (color adjectives) dependent on comparison class instead of independent*	• *Increase of notional classes: color, physical property, more specific dimension, physical state, behavior* • *Pronominal measure phrase constructions, contextual comparatives, superlatives*	*Use of multiple adjectives to describe referent unambigously (omission of size information)*

Note. italic = Findings from previous studies, bold = Findings from studies of the present thesis.

exemplar or they call also short objects *tall*. The observation that children sometimes do not relabel *big* objects as *small* when the context differs has also been claimed to be the result of so-called 'nominal interpretations' (see Table 6.4). However, it seems that children have acquired some lexical meaning of gradable adjectives, namely that these adjectives are relational in the sense that for example something is tall in relation to something that is not tall. Already at Stage 2 (see Table 6.3), children's interpretations of relative gradable adjectives seem to imply relational meaning. As argued by Sera and Smith (1987), most of the interpretations which have been claimed to be 'nominal' can be characterized as relational, too, because children have to relate objects at least to some extent to figure out

which one is tall and which one is not.

I claim that this 'relational' meaning can be generated without the notion of degrees. This claim is supported by findings from cognition studies (Ehri, 1976; Piaget & Inhelder, 1973). In these studies, children had to order objects. Children up to age 5 failed to order objects of different sizes, i.e., to construct a scale, but rather built classes of big and small objects. Similar to linguistic studies, children were able to recognize that there are big and small (or not-big) objects, but failed to establish a relation between the objects and degrees on a scale.

Hohaus et al.'s (2014) findings regarding the use of comparison constructions in spontaneous speech support the claim that initially, children do not include degrees in their semantic representation of gradable adjectives. It is possible that this interpretation is the default because there are languages without degree semantics, e.g., Motu (Beck et al., 2009).[1] According to Van Geenhoven (2006, p. 6), it is likely "that children may go through semantic stages representing particular semantic interpretations that are assigned to forms and constructions of non-target languages". Assuming that 3-year-old children's semantic representation of gradable adjectives does not include degrees, how can the aspects that are mastered in the comprehension of these adjectives and those that are not mastered be explained? I address this question in the following.

Aspects mastered and not mastered in the comprehension of gradable adjectives at Stage 3

As can be seen in Table 6.4, at age 3 children have mastered the distinction between absolute and relative gradable adjectives regarding the position of the standard: the standard for relative gradable adjectives is around the center, the standard for absolute gradable adjectives is either the maximal or minimal endpoint of the scale. I claim that this knowledge is related to the acquisition of the lexical meaning of adjectives. The 'relational' interpretations for relative gradable adjectives described above result from the emerging knowledge that properties denoted by adjectives can have opposites. This interpretation does not require that the child is able to identify an adjective as gradable.

Syrett (2007, p. 232) argues that children have different standards of comparison for relative and absolute gradable adjectives because "scales are part of [children's] conceptual repertoire", hence when identifying a novel word as an adjective, children have the expectation that the adjective can refer to a scale. However, knowledge about scales implies knowledge about degrees because scales consist of degrees. I argue that children's division of sets into big and non-big objects for instance (or big and small objects or alternatively big, small, and neither big nor small objects) is not related to the concept of scales from the outset. Instead, the division of sets is related to the concept of paths. Typically, paths encode concepts of spatial location and motion in events (Jackendoff, 1990). Syrett (2007) proposes that the structure of paths is similar to the structure of scales: both paths and

[1] Note that there are also theoretical approaches that do not make use of degrees as pointed out in Chapter 2.

scales are one-dimensional, directional, and composed of a sequence of adjacent and connected elements. A further striking similarity is that paths, like scales, can be bounded or unbounded. Paths can be bounded by a SOURCE as in (6.3a) (Jackendoff, 1983, ex. 9.9c) and/or by a GOAL as in (6.3b) (Jackendoff, 1983, ex. 9.9a).

(6.3) a. John ran from the house.
b. John ran to the house.

Syrett (2007) notes that it is not clear whether scales and paths are interchangeable concepts or whether the concepts are different but strongly connected. It has been proposed that paths belong to the major conceptual structures that are universal and innate (Jackendoff, 1990). In contrast, degrees do not seem to be universal (see Beck et al., 2009). Therefore, it may be that paths initiate scales and are the preliminary way to structure events and also properties of individuals. As a consequence, children master the distinction between absolute and relative gradable adjectives with respect to the position of the standard because they structure the sets of objects differently. It is likely that children divide the sets in the visual array by grouping objects with similar properties resulting in 'similarity' classes. This process may benefit from children's sensitivity to endpoints[2]: if there is an endpoint, this endpoint determines the boundary between the similarity classes. If there is no endpoint, the boundary between the classes must be at a different position. For instance, children may divide the sets into equal parts. As a consequence of this division of sets, antonyms of absolute and relative gradable adjectives share one degree as the standard, i.e., sets are divided exhaustively into big and non-big (or big and small) and clean and non-clean (or clean and dirty) objects. For absolute gradable adjectives, this is the target-like pattern. For relative gradable adjectives, a set of undefined objects, i.e., neither big nor small objects, is expected. Indeed, the data of the 3-year-olds suggest that they sometimes show a gap between antonyms of relative gradable adjectives, but never for absolute gradable adjectives. This finding hints at an emerging different division of sets depending on the kind of adjective. This observation does not point to a default-interpretation for gradable adjectives as either relative or absolute gradable.

In contrast to the distinction between relative and absolute gradable adjectives with regard to the standard of comparison, 3-year-old children have not mastered the distinction between relative and absolute gradable adjectives with regard to the relevance of the comparison class (see Table 6.4). How can this finding be explained by the degreeless semantic representation of gradable adjectives, which children are assumed to have at this acquisition stage?

The results of the comprehension experiment showed that children use the information of the noun to restrict the set of objects under discussion. This is in line with previous

[2] Findings from a perception study by Lakusta and DiFabrizio (2016) indicate that infants have a visual preference for endpoints in motion events. Moreover, preschool children encode GOAL paths systematically and omit SOURCE paths (Lakusta & Landau, 2005). According to Schulz (2018), the sensitivity to endpoints also guides the acquisition of verb meanings.

findings that children integrate the adjective and the noun meaning (see Table 6.3). If this were not the case, children would have selected for instance also the big soccer balls when asked for the big water balloons in an upper expansion context displayed in Figure 6.1 because the soccer balls are even bigger than the water balloons.

Figure 6.1: Upper expansion context for the RG *big*. The objects are in linear order for illustration purposes.

Hence, we can assume that children recognize that they have to include more objects, i.e., for instance water balloons and soccer balls, in case the visual array is described with a superordinate-level noun (*toys*) or fewer objects in case the visual array is described by a basic-level noun (e.g., *water balloons*). Because I assume that children divide sets by grouping objects with similar properties, it may be that more or fewer objects with a given property have the same effect on the interpretation of absolute and relative gradable adjectives. They are simply added to (or subtracted from) the already existing 'similarity' classes. However, as discussed in Chapter 5, there are only few children who never adjusted the standard for relative gradable adjectives and the adjustments of relative and absolute gradable adjectives were not entirely comparable. Therefore, again – similar to the nature of the standard – an emerging distinction between relative and absolute gradable adjectives can be observed with respect to the relevance of explicit comparison classes.

Ambiguous results

In the present comprehension experiment, 3-year-old children interpreted relative gradable adjectives as having a standard around the center of the scale. In contrast, some previous studies found interpretations of relative gradable adjectives as related only to the extremes of the scales at age 3 (e.g., Tribushinina, 2013) as summarized in the overview in Table 6.4. This finding is also in line with the approach that children have a semantic representation of gradable adjectives without degrees, but nevertheless interpret gradable adjectives as relational by grouping objects into similarity classes. It is possible that chil-

dren's preference for endpoints is just too prominent to locate the boundary between the similarity classes at a different point.

Ambiguous results were also obtained with respect to the relevance of the comparison class for the interpretation of gradable adjectives. In the present experiment the comparison class was explicit, i.e., linguistically encoded by the noun. Across test trials the noun changed, whereas the visual context did not change. The findings of the present experiment indicate that initially, the comparison class has the same non-adult-like relevance for both absolute and relative gradable adjectives. In contrast, the findings of previous studies using implicit comparison classes indicate that 3-year-olds distinguish target-like between relative and absolute gradable adjectives: they interpret absolute gradable adjectives independent of the comparison class, but relative gradable adjectives dependent on the comparison class. In these experiments the comparison class was not provided by the noun, but by the visual context. This means changes in the visual context affected the calculation of the standard for relative gradable adjectives, but not for absolute gradable adjectives. These observations raise two questions: Why do 3-year-old children not adjust the standard of comparison when the noun changes? How do implicit and explicit comparison classes differ? In other words, why do children at the same age figure out on the basis of the visual context, but not on the basis of the noun, that relative gradable adjectives are context-dependent and absolute gradable adjectives are not? This may have something to do with children's ability to identify the objects that serve as the comparison class. It is possible that changes in the visual context serve as a first cue that facilitates accessing changes in the comparison class. In addition, in previous studies 3-year-old children often had to compare only two objects from the same basic-level category instead of 12 objects from two different basic-level categories as in the present experiment. Moreover, although children at age 3 are able to use the noun to restrict the set of possible referents, they are just starting to extend adjectives to objects from different basic-level categories. It may be that at age 3, a change in linguistically encoded comparison classes does not trigger the restructuring of the similarity classes if not accompanied by changes in the visual context.

Another reason why children do not adjust the standard for relative gradable adjectives more often may be that the size differences between the objects were too small, as discussed in Chapter 5. This reason can also be related to the semantic representation of gradable adjectives lacking degrees at this third acquisition stage. Beck et al. (2009) note that Motu is supposed to be a language without degree semantics (see Chapter 2). To express comparison, e.g., *Mary is taller than Frank*, Motu uses the positive form of antonym pairs illustrated in example (2.42), repeated here as (6.4).

(6.4) Mary na lata, to Frank na kwadōgi.
Mary TOP tall, but Frank TOP short
'Mary is taller than Frank.'

Interestingly, according to Beck et al. (2009), a sentence like (6.4) is not acceptable in contexts in which Mary is 2m and Frank is 1.98m, i.e., when both are considered tall. In

contrast, in contexts in which Mary is 1.70m and Frank is 1.60m, i.e., when Mary is considered tall and Frank is considered short, (6.4) is acceptable. This observation resembles the findings from acquisition: children at Stage 3 divide sets into opposites by grouping objects with similar properties, but they do not establish an ordering within these classes by means of degrees. As a consequence, if size differences between objects are too small, they belong to the same similarity class independent of changes in the comparison class (see also Sera & Smith, 1987).

With regard to the error types, unanalyzable responses were more prominent at Stage 3 than at the later stages. This means that children left out one object in their selection of objects: for instance, they selected object 6 and 8, but not object 7. This finding indicates that children at that acquisition stage do not really order individuals according to different degrees, which is also mirrored in the results of the seriation tasks from previous cognition studies.

In summary, I propose that children around age 3 master the acquisition of gradable adjectives and the first distinction between relative and absolute gradable adjectives by specifying their lexical entries for these adjectives, but without a degree semantics for them. This way, children acquire knowledge about antonym pairs and add gradable adjectives that refer to more specific dimensions to their vocabulary. The first step in the acquisition of a target-like representation of gradable adjectives may be the child's discovery that adjectives can have opposites. This discorvery enables children to divide sets into 'similarity' classes, e.g., to categorize objects as big or small. Category formation was also found in adults when they were taught novel words with an intended relational meaning (Ryalls & Smith, 2000). This categorization "strategy" is supported by McNally (2011) who argues that one important function of adjectives is to classify individuals according to the way they manifest a given property.

Stage 4:

I assume that at acquisition stage 4, children have a semantic representation of gradable adjectives as relations between individuals and degrees. This has consequences for the compositional semantics of gradable adjectives: formally, the adjective must be combined with a *pos*-morpheme, which defines the standard of comparison. This way, the modified noun can affect the calculation of the standard degree.

Aspects mastered and not mastered in the comprehension of gradable adjectives at Stage 4

As illustrated in Table 6.5, children around age 4 use the information provided by the noun to restrict the comparison class. Moreover, they have figured out that changes in linguistically encoded comparison classes affect the interpretation of relative gradable adjectives even if the visual context does not change. In addition, they can distinguish between relative and absolute gradable adjectives because for the latter the standard of comparison

Table 6.5: Overview over the aspects mastered and not mastered in the acquisition of adjectives at Stage 4.

		Comprehension		Production	
Stage	Age	Mastered	Not mastered	Mastered	Not mastered
4	4;00	• **Distinction between relative and absolute gradable adjectives wrt to explicit comparison classes** • *Comprehension of comparatives* • *Interpretation of emotion adjectives as relative gradable*	• *Standard for relative gradable adjectives exactly at the midpoint of the scale* • **Standard for relative gradable adjectives as range of degrees: antonyms have two different standard degrees** • **Standard adjustments for relative gradable adjectives according to properties of the object distribution**	• *Correct use of degree modifiers for different gradable adjectives* • *Increase of notional classes: physical property*	

Note. italic = Findings from previous studies, bold = Findings from studies of the present thesis.

is not affected by changes in the comparison class. These aspects that are mastered in the interpretation of gradable adjectives around age 4 can be explained as follows: first, at Stage 4 children start to figure out that the opposite 'similarity' classes, e.g., the set of small objects and the set of big objects, are ordered in themselves. To order objects, children have to compare the objects to each other one by one. Similar to the task in which children have to compare one object to another object – a task they master at Stage 3 – children are now able to compare one object to multiple other objects. This way, children refine their structuring of properties via paths by constructing scales.[3] At the same time, as previous studies have shown, children understand linguistic comparison forms (Barner & Snedeker, 2008) and begin to develop ordering skillls in non-verbal tasks (Ehri, 1976).

I argued that children master the different standards of comparison for relative and absolute gradable adjectives due to the ability to recognize that sets of objects are structured

[3] This process of implementing degrees in the semantic representation of gradable adjectives may have consequences for theoretical approaches on the definition of degrees (for an overview see Solt & Gotzner, 2012). The so-called 'derived degree approach' assumes that degrees and scales are derived by grouping entities into equivalence classes and an ordering relation on these classes. In contrast, the so-called 'abstract degree approach' assumes that degrees are primitives and are not dependent on specific measurement relations between entities. The assumption that the *Semantic Complexity Hierarchy* is part of children's inherent linguistic capacity (see Section 6.2.3) favors the 'abstract degree approach'. However, it may be that children establish these abstract degrees via concrete ordering relations between entities as proposed by the 'derived degree approach'. For now, this conclusion remains speculative.

differently depending on a given property. Initially, children use their knowledge about paths and later their knowledge about scales. The lexical entry for gradable adjectives is further extended by their scale structure. But how do children acquire the relation between the scale structure and the (non-)relevance of the comparison class for the calculation of the standard? In other words, how do children know that for open scale adjectives such as *big* and *small* the standard changes if the noun changes, and that for partially closed scale adjectives such as *clean* and *dirty* the standard is fixed? This relation may correspond to semantic processing as proposed by economy principles (Kennedy, 2007; Toledo & Sassoon, 2011). According to these principles, the interpretation of a gradable adjective relative to a minimum or maximum endpoint should be preferred over a contextually-defined standard. Because the scale structure is part of the adjective's lexical meaning, children know when it is not possible to use an endpoint as the standard of comparison. As a consequence, the standard is calculated based on the context. In contrast to Stage 3, the modified noun can provide the comparison class in addition to the visual context.

Given the target-like semantic representation of gradable adjectives at Stage 4, how can the two apects that are not mastered in their interpretation be explained? The first aspect concerns the standard adjustments. In the comprehension experiment of this thesis it was found that the group of 4-year-olds often changed the standard of comparison for relative gradable adjectives when the noun changed, but not according to the distribution of objects in the comparison class. For instance, suppose that in the scenario in Figure 6.2 the standard for *big water balloons* (right picture) is the object circled with the dashed line (= object 6). When asked for *big toys* (left picture) the soccer balls should be included in the comparison class. Because they are bigger than the water balloons, the mean size of the comparison class increases. Therefore, it is expected that the standard for *big toys* is higher than the standard for *big water balloons*, e.g., the object circled with the dashed line (= object 8). Hence, fewer water balloons should have been selected. Instead, participants selected for instance the object circled with the solid line (= object 5) as the standard, and hence selected more rather than fewer water balloons.

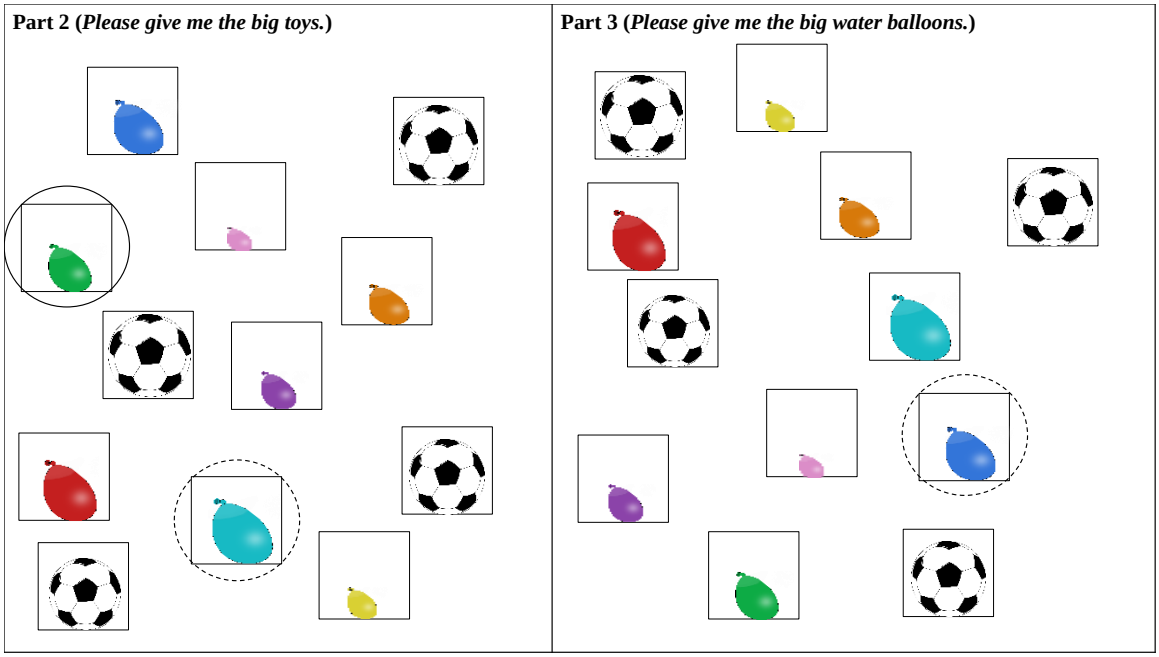


Figure 6.2: Example test trial for the RG *big* in the upper expansion context.

This finding may relate to the manipulation of the statistical properties of the comparison class, i.e., to process how bigger or smaller objects change the mean size of objects present in the visual array. As mentioned in Chapter 5, the difficulty in conceiving the distributional properties of the object array is probably more strongly connected to non-linguistic than to linguistic abilities. A relation to cognitive skills, processing capacities, or neurological maturation processes is likely. Further studies that control for these non-linguisitc abilities are necessary to better understand their consequences for linguistic abilities.

The second aspect that is not yet mastered at Stage 4 concerns the definition of the standard of relative gradable adjectives as a range of degrees (see Solt, 2011). This definition accounts for the so-called 'borderline-cases', which are one characteristic of vague interpretations for relative gradable adjectives. There is a set of objects that cannot be clearly judged as big or small for example. The group results of the present comprehension experiment suggest that 4- and 5-year old children locate the standard exactly at the midpoint of the scale, similar to the findings by Tribushinina (2013) (see Table 6.5). However, the individual analysis of the children's responses in the present comprehension experiment conducted in Chapter 5 yielded a more informative result. This analysis showed a 'gap' between the objects 4-year-olds judged big and the objects 4-year-olds judged small. However, I claimed in Chapter 5 that they calculate two different standards for *big* and *small* rather than treating the standard as a range of degrees because sometimes they judged the same object both big and small. In this case, it can be assumed that the two standards are reversed because children do not respect that the standard for *big* must be always higher than the standard for *small*. The 'gap' and the standard reversals found in the individual analysis can be explained by children's relational interpretation of relative

gradable adjectives. On the one hand, it is possible for 4-year-olds to label the same object *small* and *big* depending on the context. If the visual context does not change as in the present experiment, this interpretation can result in reversed standards for the positive and negative antonym. On the other hand, there are objects that 4-year-olds label neither *small* nor *big*, which results in the expected 'gap'.

In summary, around age 4, children's interpretation of adjectives as gradable or relational results from including degrees in the semantic representation. Besides the lexical specification of gradable adjectives by taking the scale structure into account, at Stage 4 the compositional semantics of gradable adjectives in form of the *pos*-morpheme contributes to the target-like interpretation of gradable adjectives. The standard of comparison is calculated with respect to a comparison class, which can be linguistically encoded by the noun. The adjective's scale structure can be used as a cue for whether the value of the comparison class is relevant for the calculation of the standard or not.

Stage 5:
Children have a target-like semantic representation of gradable and non-gradable adjectives around age 4; at Stage 5 two further developments in their interpretation can be observed.

Table 6.6: Overview over the aspects mastered and not mastered in the acquisition of adjectives at Stage 5.

		Comprehension		Production	
Stage	Age	Mastered	Non-mastered	Mastered	Non-mastered
5	5;00	• **Standard for relative gradable adjectives as range of degrees** • *Sensitivity to object class (world knowledge) when calculating the standard* • **Standard adjustments for relative gradable adjectives according to properties of the object distribution**			

Note. italic = Findings from previous studies, bold = Findings from studies of the present thesis.

First, the findings from the present comprehension experiment summarized in Table 6.6 indicate that at age 5, children's standard for relative gradable adjectives is defined as a range of degrees. The judgements for the same objects as both big and small decreased, while a 'gap' was still present between big and small objects, i.e., objects that were judged

neither as big nor as small. How do children come to an interpretation of relative gradable adjectives that leaves some objects in a series undefined? This interpretation may be related to children's developing flexibility in the calculation of the standard. Tribushinina (2013) proposes that from age 5, children start to use different kinds of information such as the visually presented comparison class and world-knowledge about the comparison class when they calculate the standard of comparison. This additional information may prompt children to locate the standard not necessarily at the exact midpoint of the scale.

The second aspect in the interpretation of relative gradable adjectives developing at Stage 5 may also interact with the nature of the standard described above. At Stage 5, the number of standard adjustments in the unexpected direction given a specific distribution of objects decreases. Thus, it seems that children are able to integrate additional objects in the comparison class correctly and adjust the standard of comparison accordingly. Interestingly, there was a tendency, also in adults, to adjust the standard more often when the size of the additional objects coincided with the polarity of the adjective (see Chapter 5). That means the standard was more often adjusted for *big* in contexts in which the additional objects were bigger than the original objects (see Figure 6.2) and it was more often adjusted for *small* in contexts in which the additional objects were smaller than the original objects. This so-called 'semantic congruity effect' was also found in other studies with adults (Hansen & Chemla, 2017; Ryalls & Smith, 2000). It seems that people base their interpretation of relative gradable adjectives on examples for which there is no conflict between the adjective and the cues from the visual context or from world-knowledge. Nevertheless, it is open how exactly the development regarding the standard of comparison for relative gradable adjectives can be explained.

To conclude, in this section I proposed a developmental path for the acquisition of the semantics of adjectives. Around age 2, children know that adjectives refer to properties of objects rather than to the object category. That is, adjectives are taken to describe a property of an entity or to specify the set of possible referents in production and comprehension. I argued that children start with a semantic representation of all kinds of adjectives as denoting properties of individuals (type $\langle e, t \rangle$) (see Hohaus et al., 2014). Around age 4, this semantic representation is enriched by degrees. As a result, children are able to interpret adjectives as gradable. Between age 2 and 4, the distinction between different kinds of adjectives may emerge via the conception of different kinds of paths and the construction of similarity classes, which result in different categorization patterns. As of age 5, the interpretation of gradable adjectives develops further (see Hohaus et al., 2014; Nelson, 1976; Tribushinina, 2013). It is open at which age children understand that adjectives can be non-intersective (e.g., *former*) and at which age children produce members of this adjective class. The production study of this thesis revealed that non-intersective adjectives are acquired later than other adjective classes. This pattern may also hold for their comprehension.

6.2.2 Production-comprehension asymmetry in the acquisition of adjectives

In this section, I address the question of why adjectives are produced before they are interpreted target-like. In Section 6.2.1, I argued that at age 3, children do not have the target-like semantics for some adjectives. In particular, I proposed that degrees are not part of children's semantic representation of gradable adjectives yet. Hence, the semantic representation of gradable adjectives does not differ from the representation of non-gradable adjectives. This acquisition stage is based on findings regarding the comprehension of adjectives. At first sight, this result seems to be in conflict with the findings for the production of adjectives showing that gradable adjectives are produced in spontaneous speech before the age of 3 and more importantly, different adjective classes emerge in an orderly fashion predicted by semantic properties.

This asymmetry raises the question of how the order of acquisition in spontaneous speech can be determined by semantic properties although children do not seem to interpret adjectives according to these semantic properties. Recall that hypothesis H1 stated that the order of acquisition of adjectives is determined by their semantic complexity. In Chapter 2, I proposed that semantic complexity should be defined as a combination of the adjective's length of description, its type complexity, and lexical properties that contribute to the adjective's meaning. In the following, I suggest that lexical semantic properties may be more prominent than compositional semantic properties at the beginning of the acquisition process and can account for the acquisition order in spontaneous speech. As a consequence, the order of acquisition is compatible with the *Semantic Complexity Hierarchy* although children's semantic representation is not target-like as argued in Section 6.2.1.
Table 6.7 summarizes the components that yield the adjective classes' semantic complexity.

Table 6.7: Semantic type and lexical properties of adjective classes 1 to 4.

Class	Semantic type	Lexical properties
1 'BLUE'	$\langle e,t\rangle$	/
2 'CLEAN', 'BIG'	$\langle d,\langle e,t\rangle\rangle$	scalar structure [a]
3 'HEALTHY', 'BEAUTIFUL'	$\langle d,\langle e,t\rangle\rangle$	scalar structure, abstract scales, multiple dimensions, judge/experiencer
4 'FORMER'	$\langle\langle s,\langle e,t\rangle\rangle\langle e,t\rangle\rangle$	?

[a] As illustrated in Section 6.2.1, scalar structure is preceded by path structure.

Table 6.7 shows that subsective/intersective gradable dimensional adjectives (Class 2 'CLEAN', 'BIG') include more lexical properties than subsective/intersective non-gradable adjectives (Class 1 'BLUE') and fewer lexical properties than subsective/intersective gradable evaluative adjectives (Class 3 'HEALTHY', 'BEAUTIFUL'). It was found in the spontaneous speech data analysis that adjectives in combination with nouns emerge descriptively in the following order: Class 1 before Class 2 before Class 3, as predicted by the *Semantic*

Complexity Hierarchy. However, this order can also be predicted by the number of lexical semantic properties while the semantic type is the same for all classes, namely $\langle e, t \rangle$. It is likely that these lexical properties are triggered by notional classes (COLOR, DIMENSION, BEHAVIOR etc.). It is important to note that the production-comprehension asymmetry does not reject hypothesis H1. The acquisition order of adjectives in spontaneous speech is determined by their semantic compelxity, i.e., by the *Semantic Complexity Hierarchy*, but it seems that young children can access or activate only parts of it, namely the lexical semantic properties.

This partial access may also be an explanation for the observation that non-intersective adjectives such as *former* or *alleged* were not produced until the age of 3. For this adjective class it is unclear whether specific lexical properties contribute to the adjectives' meaning, but, crucially, adjectives from this class cannot be interpreted as intersective. Hence, if the intersective property-denoting interpretation of adjectives is the default, adjectives like *former* or *alleged* cannot be derived.

Alternatively to this approach, children may have a target-like semantic representation for all adjectives from the outset. However, this representation is only mirrored in the production, but not in the comprehension of adjectives. Because all compositional and lexical semantic properties of adjectives are available to the child, these properties determine the specific order in which adjectives emerge in spontaneous speech. The delayed comprehension may be explained either methodologically or psycholinguistically (see Ünal & Papafragou, 2016, for an overview). According to methodological explanations, production-comprehension asymmetries should be seen as an epiphenomenon resulting from too demanding tasks. The asymmetry should disappear when memory or metalinguistic demands are lowered. Another methodological reason for different outcomes in production and comprehension may be that different methods are used to measure production and comprehension (Ünal & Papafragou, 2016). To find evidence for or against a methodological explanation regarding the production-comprehension asymmetry in the acquisition of adjectives, further experiments in addition to the two studies of this thesis are necessary.

Psycholinguistic explanations ascribe production-comprehension asymmetries to different processes involved in comprehending and producing utterances. The delayed comprehension may be attributed to either lacking processing abilities or to lacking pragmatic or metalinguistic abilities (Ünal & Papafragou, 2016). Thus, although children have a target-like semantic representation, it is possible that they cannot retrieve all the information relevant for the interpretation of an utterance in comprehension experiments. Looking at the interpretation of focus particles, Müller (2012) claims that non-target-like responses are more likely in situations in which the position of the focus particle is in conflict with the typical information structure of the sentence. In these conflict situations, the child focuses on only one information available which results in a non-target-like response. However, for younger children it is possible that non-target-like responses result from a non-target-like semantics of focus particles. With regard to the interpretation of gradable adjectives,

it is possible that for 3-year-olds the conflict in comprehension experiments arises from their knowledge about gradability applied to adjectives with different lexical properties, i.e., the different scale structures associated with relative and absolute gradable adjectives. However, as mentioned by Müller (2012), it is also feasible that at this age, children have not acquired all semantic aspects necessary for the interpretation. Therefore and for the reasons explained in Section 6.2.1, I assume that children's access to the lexical properties in the *Semantic Complexity Hierarchy* such as scalar structure is prior to the activation of semantic properties such as the notion of degrees. A final answer to the question of the production-comprehension asymmetry requires further research.

6.2.3 Innate knowledge and positive evidence for the acquisition of adjectives

Based on the acquisition path developed in Section 6.2.1, this section addresses the question of how children can reach the next stage in the time course of adjective aquisition. In principle, two scenarios are possible. On the one hand, the development can be driven by the input. In this case the child derives abstract linguistic representations by generalizing across specific cases that occur in the linguistic environment. This generalization can be achieved via statistical learning or via extralinguistic conceptual cognition such as associate learning. On the other hand, the development can be driven by inherent linguistic knowledge. In this case the child is equipped with abstract linguistic representations and uses the input to identify their realizations in the target language. In both scenarios the input plays an important role for the acquisition of language, thus it is likely to find effects of statistical-distributional patterns under both approaches (Lidz & Gagliardi, 2015). However, children are able to produce and comprehend sentences that they have never heard before and to distinguish possible from impossible sentences, so that they do not adopt ungrammatical features although they may be present in the input. This 'poverty of the stimulus' can be seen as an argument that children's linguistic knowledge cannot be based on input alone. Rather, children's linguistic knowledge is derived from their inherent linguistic capacity, the Universal Grammar (UG). Lidz and Gagliardi (2015) assume that UG must be complemented by mechanisms that allow the learner to extract the information from the input that is critical for constructing the target grammar. As a consequence, learners must be biased to attend to certain information. The input can be filtered by UG, i.e., childrens' expectations associated with the hypotheses regarding a particular grammar, processing mechanisms, prior knowledge, or conceptual representations (E. V. Clark, 2004; Lidz & Gagliardi, 2015; Mandler, 2012). Based on the filtered input, children build a so-called 'perceptual intake representation' (Lidz & Gagliardi, 2015) that is not identical to the adult representation in the sense that it does not contain every information present in the adult representation. According to Lidz and Gagliardi (2015), this perceptual intake representation is then compared to the features of the grammar predicted by UG. This comparison allows inferences about the grammatical features crucial for the sentence that

the child has heard and updates the grammatical knowledge that is used for the analysis of subsequent sentences.

During the time course of acquisition children extract different information from the input (Lidz & Gagliardi, 2015) and attend to different cues (*emergentist coalition model*, Hirsh-Pasek, Golinkoff, and Hollich (2000)). According to Hirsh-Pasek et al. (2000), the cues for word learning in the input are available to the child from the beginning, but weighted differently over time depending on how easily the cue can be detected by the child. In addition, Kauschke (2007) notes that the cues for word learning may be of different importance depending on the word class. Cues can be provided at different levels: the language learner can receive linguistic cues (phonological, syntactic, semantic, pragmatic), distributional cues, or conceptual cues. In the following, I propose what kind of knowledge may be predisposed and what kind of positive evidence the child may need to further specify the innate knowledge and to reach the target-state in the acquisition of adjectives in the exposure language, in this case German. Note that for the time being, I can provide evidence for the existence of potentially informative structures in the input[4], but not to which amount this information is available. It is open how many of these structures the child would have to be exposed to and whether the child can use the information provided in the input.

I suggest that the child enters the acquisition process at Stage 1 with linguistic and cognitive predispositions. The *Semantic Complexity Hierarchy* is a possible candidate to be part of the child's inherent linguistic capacity. It is possible that the *Semantic Complexity Hierarchy* is not specific to adjectives, and may be applied to other word classes such as nouns. Although nouns do not exactly behave like adjectives, they can be gradable, too (Morzycki, 2014; Sassoon, 2017). Sapir (1944) also notes that gradability is in principle cross-categorial. Besides the linguistic predisposition, children have a cognitive predisposition to categorize their environment. The concept of paths and the ability to make comparisons are possible candidates to be part of children's cognitive abilities. Now, given the linguistic and cognitive predispositions, how can positive evidence contribute to establishing the word class ADJECTIVE and the differences between subcategories of adjectives?

To arrive at Stage 2, the child must acquire (a) that adjectives refer to object properties rather than to object categories and (b) that adjectives can modify nouns. Regarding (a), it is possible that conceptual cues are more accessible to the two-year-old child than linguistic or distributional cues. To acquire (b), distributional and syntactic cues may be informative. The child may detect that an adjective occurs between a determiner and a noun and is inflected. Because in this attributive position adjectives restrict the set of possible referents, the adjective may be stressed (phonological cue) to contrast the respective property with other properties. The child may also recognize that the same word can occur together with a noun and a copula verb and is in this structure uninflected. Both structures can be found in the input of Leo at age 2;00:

[4] The data are taken from the "Leo" corpus described in Chapter 4.

(6.5) a. *MUT: die Kerze ist jetzt auch rot, genau.
*MUT: the candle is now also red right
'*MUT: the candle is red now, too, right.'

b. *MUT: das hat Punkte, genau, das rote Halstuch.
*MUT: this has dots, right, the red scarf
'*MUT: this has dots, right, the red scarf.'

To arrive at Stage 3, the learner's task is to distinguish between different adjective classes. To differentiate non-gradable and gradable adjectives, pragmatic cues may be available. Sandhofer and Smith (2001) report that parents use color, i.e. non-gradable, and size, i.e. gradable, adjectives in different contexts: color adjectives are used in single object contexts in questions such as *What color is this?*, hence asking the child to produce the color word. In contrast, size adjectives are used in multiple object contexts in questions such as *Where's the big one?*, hence the child must comprehend the adjective to answer the question. This may be a pragmatic cue to differences between adjective classes because the adjectives are used in different situations. A similar difference is also found in German (see examples (6.6a) and (6.6b)). Note that it is not clear whether example (6.6b) is uttered in a context with more than one rabbit.

(6.6) a. *MUT: was ist das für (ei)ne Farbe hier, xxx ?
*MUT: what is this for a color here
'*MUT: what color is this?'

b. *MUT: wo is(t) der große Hase?
*MUT: where is the big rabbit
'*MUT: where is the big rabbit?'

Semantic and distributional cues may serve to detect that an adjective is gradable. A semantic cue may be the use of antonym pairs (see Tribushinina, van den Bergh, et al., 2013) indicating that properties denoted by adjectives can have opposites and can be used to compare objects like in example (6.7). Stress on the adjectives can be a phonological cue to strengthen the contrast between antonym pairs.

(6.7) *MUT: +< ja, das sind die kleinen Katzen und das ist die große Katze.
*MUT: yes, these are the small cats and this is the big cat
'*MUT: yes, these are the small cats and this is the big cat.'

Marking an adjective with comparative morphology as in example (6.8) or the use of some degree adverbs such as *very* (see example (6.9a)) or *too* (see example (6.9b)) is only possible for gradable adjectives. Hence, this is a distributional cue to distinguish between gradable and non-gradable adjectives.

(6.8) *VAT: und welcher von den beiden ist größer, sag mal.
*VAT: and which of the both is bigger, tell
'*VAT: tell me, which of the two is bigger.'

(6.9) a. *MUT: +< die sind alle sehr sehr groß, Leo.
*MUT: these are all very very big, Leo
'*MUT: all of them are very big.'

b. *MUT: hm, da habe ich da den kleinen Verdacht, dass der große
*MUT: hm, there have I there the small suspicion, that the big
Bagger zu groß ist, für den kleinen Tunnel.
digger too big is, for the small tunnel
'*MUT: I suspect that the digger is too big, for the small tunnel.'

The distinction between absolute and relative gradable adjectives may also be discovered by the distribution of degree adverbs. As shown in Chapter 2, degree adverbs like *almost, completely, slightly* can modify absolute gradable adjectives, but not relative gradable adjectives. Syrett (2007) showed that English-speaking 3-year-old children are sensitive to the distribution of degree adverbs when assigning an absolute gradable or relative gradable interpretation to a novel adjective. However, in German at least *ein bisschen* ('slightly') does not modify unambiguously absolute gradable adjectives. In example (6.10a) *ein bisschen* modifies the verb. In example (6.10b) *ein bisschen* quantifies over a noun.

(6.10) a. *MUT: jetzt mach mal die Tür wieder zu und lass das Auto noch ein
*MUT: now make the door again shut and let the car a
bisschen fahren, Leo.
bit drive, Leo
'*MUT: now close the door and let the car drive.'

b. *MUT: Leo, [//] magst du noch (ei)n bisschen Müsli?
*MUT: Leo, want you again a bit muesli
*MUT: do you want some muesli, Leo?'

c. *VAT: der ist eigentlich ein bisschen klein, der Tunnel, hm, für den Laster,
*VAT: this is actually a bit small, the tunnel, hm, for the truck,
genau genommen.
exact taken
'*VAT: In fact, the tunnel is a bit too small for the truck.'

In example (6.10c) *ein bisschen* modifies a relative gradable adjective but the meaning is different compared to absolute gradable adjectives. The intended meaning in (6.10c) is that the tunnel is *too* small for the truck. Therefore, in German not all degree adverbs may be a reliable cue for the distinction between relative and absolute gradable adjectives (see Tribushinina, 2017, for Dutch and Russian).

Distributional cues for the class of evaluative adjectives may be for instance the co-occurrence with *finden* ('to find') as in example (6.11), which indicates that a judge can be involved, or the co-occurrence with *außer* ('except for'), which indicates that the adjective is associated with more than one dimension. For *außer* no instance in combination with an adjective was found.

(6.11) *VAT: das find(e) ich gar nich(t) lustig.
*VAT: this find I not funny
'*VAT: I do not find it funny at all.'

The most important aspect in which Stage 3 and Stage 4 differ is the implementation of degrees in the learner's semantic representation. How can the learner detect that the target language makes use of degrees? According to Hohaus et al. (2014), the degree argument of an adjective is overtly realized in degree questions as in example (6.12a) or by overt measure phrases as in (6.12b).

(6.12) a. *MUT: wie groß war der denn?
*MUT: how big was this
'*MUT: how big was this one?'

b. *VAT: also, ja, der is(t) schon hundert fuffzig Jahre alt.
*VAT: this is already hundred fifty years old
'*VAT: this one is already one hundred and fifty years old.'

In summary, during the time course of acquisition the input provides different cues for the specification of adjectives that is predisposed by the *Semantic Complexity Hierarchy*. A more detailed analysis of the input is necessary to determine whether the cues proposed here are really unambiguous and informative to allow inferences regarding possible and impossible adjective structures and interpretations.

6.3 Implications of the acquisition path

The acquisition path proposed in this chapter concerns the compositional and lexical semantic properties of adjectives and is based on experimental findings from monolingual German-speaking children. In this section, I briefly discuss the implications of this acquisition path for other classification systems of adjectives, for the acquisition of adjectives across languages, and for the acquisition of other modification structures.

6.3.1 Different classification systems of adjectives

As introduced in Chapter 2, there are different systems to classify adjectives. The system of notional classes focuses on the concepts in the mental lexicon expressed by adjectives. The semantic classes used in the *Semantic Complexity Hierarchy* focus on the adjectives' function in nominal modification structures. It is open whether these classification systems are related and how a possible relation may look like.

Ordering restrictions of adjectives as in example (6.13a) have been described by using notional classes as in (6.13b).

(6.13) a. Ein fantastischer, großer, heißer, schwarzer Kaffee
'A fantastic, tall, hot, black coffee'

b. Det VALUE > DIMENSION > PHYSICAL PROPERTY (Temperature) > COLOR N

In addition, notional classes have been employed to describe the acquisition order of adjectives. The explanations for ordering restrictions and order of acqusition have a semantic 'flavor'. For instance, Blackwell (2005) concludes that adjectives with a noun-dependent meaning are acquired later than adjectives with a noun-independent meaning. Regarding the ordering restrictions of multiple prenominal adjectives, it has beem claimed that adjectives denoting properties inherent to the referent of the noun, i.e., objective and physical properties, appear closer to the noun than adjectives denoting properties not inherent to the referent of the noun, i.e., subjective and less or non-physical properties (see Alexiadou et al., 2007, for an overview). Therefore, a relation between conceptual and semantic properties of adjectives is very likely. One could go so far as to say that notional classes follow from formal semantic properties (see McNally, 2016). As shown in Chapter 4, it is possible to apply the semantic classes in the *Semantic Complexity Hierarchy* to the notional classes but the correspondence between the two classifications is only a partial one. Therefore, I assume that both semantic and notional properties are necessary to construct a lexical entry for an adjective. Regarding the acquisition path, it is possible that notional properties trigger the lexical semantic properties that are included in the *Semantic Complexity Hierarchy*, which I claimed to be part of children's inherent linguistic capacity. Then, compositional semantic properties determine in which modification structures adjectives with specific notional and lexical semantic properties are licensed. Hence, notional properties may trigger semantic properties but the latter provide a more linguistic explanation of why adjectives are acquired in a specific order and probably also why they occur in a specific order in prenominal position.

6.3.2 Acquisition of adjectives cross-linguistically

Research on adjectives across languages mentions typological differences with respect to the adjective class (Dixon, 1982, 2004). Dixon (2004) claims that an adjective class exists in every language, but that in some languages this class is a closed class and restricted to only few adjectives. Moreover, in some languages the difference between adjectives and nouns and verbs, respectively, is subtle (Dixon, 2004). Typological differences with regard to conceptual adjective classes and with regard to semantic features (see Beck et al., 2009) raise the question of how universal the *Semantic Complexity Hierarchy* can be and how semantic complexity determines the acquisition of adjectives cross-linguistically.

The claim that the *Semantic Complexity Hierarchy* is part of children's inherent linguistic capacity means for languages with the same adjective classes like German that the acquisition order should be similar to the German acquisition order found in the studies of this thesis.[5] During the acquisition process, the different compositional and lexical

[5] This expectation is in line with the similar acquisition order of notional adjective classes across languages (Tribushinina et al., 2014).

semantic properties implemented in the *Semantic Complexity Hierarchy* are activated by positive evidence. Depending on language-specific properties the activation of the features may stop at different levels of the hierarchy. For example, for languages that are claimed to not have degree semantics, one can speculate that speakers may access semantic features of the hierarchy until Class 1 (subsective/intersective, non-gradable). Nevertheless, it should be possible to express concepts like size or evaluation in these languages, but probably not in structures which require a degree semantics. One can hypothesize that these languages may have partial access to the lexical properties beyond the Class 1-level as I claimed for early stages in the acquisition of adjectives in German (see Section 6.2.2).

6.3.3 Acquisition of other nominal modification structures

This thesis focuses on the semantics adjectives. The findings of the studies presented in Chapter 4 and 5 point to an influence of semantic complexity on the acquisition of adjectives, in particular when adjectives are used as modifiers of nominal constituents, i.e., in attributive (e.g., *the small elephant*) and predicative structures (e.g., *the elephant is small*). Modifiers include also relative clauses as in (6.14a), prepositional phrases (PP) as in (6.14b), and focus particles as in (6.14c).

(6.14) a. The elephant$_{(,)}$ **who can fly**$_{(,)}$ is called Dumbo.
b. The elephant **with the large ears** can fly.
c. The elephant and the mouse are gray. **Only** the elephant has a trunk.

How the operationalization of semantic complexity proposed in this thesis applies to these nominal modification structures and whether the acquisition of these structures is influenced by semantic complexity, is open. To evaluate how semantically complex these structures are, a detailed semantic analysis of the three structures is necessary, which is not possible within the scope of this thesis. As a starting point, I consider similarities of and differences between adjectives and relative clauses, PPs, and focus particles.

Few studies regarding the acquisition of other modifiers take semantic properties into account. Trabandt (2017) investigated the acquisition of restrictive and appositive readings of relative clauses such as (6.14a). In the restrictive reading, the relative clause restricts the set of possible referents. In the appositive reading, the relative clause specifies a referent that is already established. According to Trabandt (2017), appositive relative clauses are more complex than restrictive ones because they cannot combine with the noun via predicate modification, i.e., set intersection, but require a specific mode of composition. Her findings suggest that appositive readings of relative clauses are acquired later than restrictive readings. Thus, together with the present findings for adjectives, it seems that combining two expressions of type $\langle e, t\rangle$ via predicate modification is easier than combining expressions of other semantic types.

Predicate modification also operates in noun phrases modified by PPs as in (6.14b). Nevertheless, Hurewitz et al. (2000) found that 4- to 5-year-old children have difficulty

interpreting the PP *on the napkin* as a description of the location in sentences such as *Put the frog on the napkin into the box* in order to find the correct referent. Even in contexts in which two frogs were present, the PP was not used to restrict the set of possible referents. Previous findings on adjectives showed that children use the prenominal adjective to find the correct referent with 2;06 years. For PPs, factors other than semantic complexity may play a role such as the processing of modifiers in prenominal *versus* postominal position, which make sentences such as *Put the frog on the napkin into the box* locally ambiguous.

Focus particles or focus-sensitive operators modify the focused element in an utterance. This can be the noun as in example (6.14c). The class of focus particles is diverse, semantically they all establish a contrast between the alternative set created by the focus and the focused element. The lexical meaning of the focus particle determines the relation between the alternative set and the focused element. In a sentence with the restrictive focus particle *only* as in (6.14c) the property of having a trunk must be true for the focused element ('the elephant') and must not be true for the alternative set, e.g., 'the mouse'. The relation between the alternative set and the focused element is different for other focus particles such as additive *also*. There may be similarities between the class of focus particles and the class of gradable adjectives. The class of focus particles shares the same semantics and the class of gradable adjectives shares the same semantics. For focus particles, this is the contrast between the alternative set created by the focus and the focused element. For gradable adjectives, this is the relation between individuals and degrees. For both, lexical semantic properties are responsible for differences in meaning. For focus particles, being restrictive or additive is part of the lexical meaning. For gradable adjectives, being a closed-scale, i.e., absolute gradable, or an open-scale, i.e., relative gradable adjective, is part of the lexical meaning. With respect to gradable adjectives, I argued that they are equally complex, and hence should be acquired at the same age. Given the similarities between adjectives and focus particles one may expect that the different focus particles are equally complex, too, and acquired at the same age.

To conclude, this chapter showed that the acquisition of adjectives in production and comprehension is influenced by the adjective classes' semantic complexity as expressed in the *Semantic Complexity Hierarchy*. Notional properties of adjectives and positive evidence in the input can serve as a trigger for establishing the compositional and lexical semantic properties in the *Semantic Complexity Hierarchy*. Further research should examine the role of semantic complexity for the acquisition of other nominal modification structures more closely. The present chapter also revealed that it is crucial to take both production and comprehension data into account because asymmetries can occur between these domains.

Chapter 7

Conclusion and outlook

This thesis investigated the acquisition of compositional and lexical semantic properties of adjectives in German-speaking children between the age of two and five years. The semantic properties I discussed focused on entailment and gradability properties of adjectives. According to formal semantic approaches, there are intersective and non-intersective adjectives, subsective and non-subsective adjectives as well as gradable and non-gradable adjectives. These properties concern the compositional mechanisms involved in nominal modification, i.e., the combination of adjectives and nouns. In addition, adjectives differ regarding lexical semantic properties that contribute to the adjectives' meaning. Differences in the adjectives' scale structure have led to the theoretical assumption that gradable adjectives should be distinguished into relative and absolute gradable adjectives. In addition, meaning components such as multidimensionality or subjectivity have led to the distinction between dimensional and evaluative gradable adjectives. In Chapter 2, I argued for an integration of these properties. This integration of entailment and gradability properties results in six adjective classes constituting the *Semantic Complexity Hierarchy*. Assuming that these adjective classes differ in semantic complexity, I proposed a definition of semantic complexity that takes into account the adjectives' length of description, their type complexity, and lexical properties that contribute to the adjectives' meaning.

Previous research addressing the question of how children acquire the compositional and lexical semantic properties of adjectives is scarce. Nevertheless, previous studies revealed important insights regarding the emergence of adjectives in children's early vocabularies and the specification of the adjective class with respect to notional semantic properties such as COLOR, PHYSICAL PROPERTY, MENTAL STATE etc. In addition, previous studies investigated the development of the structures in which adjectives can be used. With regard to compositional and lexical semantic properties previous research focused on the acquisition of relative gradable dimensional adjectives. Few studies also considered the acquisition of other adjective classes such as absolute gradable dimensional adjectives. These studies provide first evidence that differences between adjectives in notional as well as in compositional and lexical semantic properties are mirrored in the acquisition of adjectives. However, these studies leave open how specific acquisition orders can be explained. Vari-

ous factors have been proposed to determine these orders, among them notional semantic properties of adjectives or properties of the input such as frequency.

On the basis of the theoretical assumptions about semantic complexity, I stated that semantic complexity determines the acquisition order of adjectives (H1). H1 predicts that more complex adjectives are not acquired before less complex adjectives, and that adjectives of the same complexity are acquired simultaneously. In addition, I stated that the frequency of adjective use in the input does not completely determine the order of acquisition (H2) because previous findings regarding the role of the input for the acquisition of adjectives are inconclusive. It is beyond question that the input plays a role in lexical acquisition, but it is unclear whether it is the major factor with respect to the acquisition of entailment and gradability properties. Hence, a relation between adjective use by the children and their caregivers is likely, but the age of acquisition is not necessarily predicted by input frequency.

The two empirical studies I conducted provide first evidence for both hypotheses H1 and H2. The spontaneous speech data analysis investigated the influence of semantic complexity and the influence of adjective frequency in the input on the acquisition of adjectives. With regard to the influence of semantic complexity, it was examined whether the *Semantic Complexity Hierarchy* is reflected in the mean age of acquisition for the adjective classes and in their growth patterns, i.e., the age at which the number of types per class increased most. The findings indicate that adjective classes of the same complexity differ neither in their mean age of acquisition nor with respect to the age at which they increased most. Adjectives of different complexity are acquired in the following order: first, on average at 2;04 years, subsective/intersective non-gradable adjectives are acquired as (attributive) modifiers of nouns followed by subsective/intersective gradable dimensional adjectives at 2;06 years and by subsective/intersective gradable evaluative adjectives at 2;08 years. Non-subsective/non-intersective adjectives are the last class acquired because they are not produced until age 2;11, the end of the time period investigated. For each class, three months before the mean age of acquisition the number of types increases most. These sequences support hypothesis H1 because more complex adjective classes are not acquired before less complex adjective classes, and adjective classes of the same complexity are acquired simultaneously.

Regarding the influence of the input on the acquisition of semantic properties, it was examined whether adjective frequency in the speech of the child and his caregivers are correlated. As predicted by H2, the analysis revealed a positive correlation pointing to a relation between the adjective use by the child and his caregivers: the more frequent an adjective is used by the caregivers, the more frequent it is produced by the child and *vice versa*. In addition, it was examined whether the age of acquisition and the frequency of adjective use in the input are correlated. The analysis revealed a negative correlation between age of acquisition and adjective frequency in the input: the more ferquent an adjective is used by the caregivers, the earlier it is acquired by the child. This finding is not predicted by H2. However, an analysis of specific adjectives revealed that adjectives that are acquired late are nevertheless present in the input from the beginning of adjective acquisition. More-

over, adjectives with a similar frequency in the input were not acquired simultaneously. This finding is predicted by H2 because the age of acquisition is not determined by the input frequency.

The comprehension experiment, a forced picture-choice task, focused on two specific classes in the *Semantic Complexity Hierarchy*: absolute and relative gradable dimensional adjectives. These two classes are of the same semantic complexity because they are both gradable, but they have different scale structures. As a result, they must be interpreted differently due to lexical semantic properties. It is predicted by H1 that absolute and relative gradable dimensional adjectives should be interpreted target-like at the same age. In particular, it was investigated whether children calculate different standards of comparison for relative and absolute gradable adjectives, and whether they interpret relative gradable adjectives, but not absolute gradable adjectives, dependent on the comparison class. Furthermore, it was assessed whether interpretation patterns change with age. The results indicate that as of age 3, children distinguish between relative and absolute gradable adjectives with regard to the standard of comparison: their standard for relative gradable adjectives is located around the center of the scale, their standard for absolute gradable adjectives is located at the minimal or maximal endpoint of the scale. However, with respect to the relevance of the comparison class, for 3-year-old children, unlike for 4- and 5-year-olds, changes in the noun, i.e., in the explicit comparison class, led to non-adult-like responses regarding both relative and absolute gradable adjectives. These findings suggest that 3-year-olds have not yet fully mastered the interpretation of absolute and relative gradable dimensional adjectives. As predicted by H1, 4-year-olds interpret both relative and absolute gradable adjectives target-like.

On the basis of the empirical findings of the comprehension experiment, I proposed an acquisition path stating that children enter the acquisition process with inherent linguistic knowledge, the *Semantic Complexity Hierarchy*, and cognitive abilities to categorize their environment via paths. I argued that initially, children interpret all adjectives as denoting properties of individuals. In other words, I claim that at this acquisition stage children do not have a target-like semantic representation of gradable adjectives as denoting relations between individuals and degrees, but that children have inherent knowledge about paths to structure properties of individuals. This way, and by making use of their emerging knowledge about antonym pairs children construct similarity classes. As a result, children at this acquisition stage are able to calculate a target-like standard of comparison for gradable adjectives without having a target-like semantic representation of gradable adjectives. As soon as degrees are included in the semantic representation of gradable adjectives, children distinguish the class of gradable adjectives from the class of non-gradable adjectives and their interpretation of different kinds of gradable adjectives becomes target-like.

At first sight, this acquisition path seems to conflict with the finding that between age 2 and 3 the adjective classes emerge in spontaneous speech in the order predicted by their semantic properties. Strictly speaking, this prediction presupposes a target-like semantic representation. However, I argued that it is possible for children to access the *Semantic*

Complexity Hierarchy only partially. At this point, factors in addition to compositional and lexical semantic properties should be considered as triggers to specify children's initial representation of adjectives. Notional properties can help to access lexical semantic properties implemented in the *Semantic Complexity Hierarchy* such as scale structure or subjective meanings. Even if initially children can access only the lexical semantic properties of the *Semantic Complexity Hierarchy*, the same predictions follow for the order of acquisition. Therefore, the findings from the spontaneous speech data analysis can be explained without contradicting the proposed acquisition path. Cues from the input may additionally signal differences between the adjective classes in the *Semantic Complexity Hierarchy*.

The findings of this thesis are a first step to understand which properties of adjectives play a role for their acquisition. However, further research is necessary to support the proposal that semantic complexity determines the acquisition of adjectives. With regard to the production of adjectives, data from more children are needed for a more elaborated statistical analysis and in order to determine whether the results of the present spontaneous speech data analysis generalize across learners. Data from different languages can reveal whether the present findings generalize across the acquisition of adjectives in different languages. In addition, cross-linguistic studies investigating the production and interpretation of gradable adjectives in languages for which it is assumed that they are lacking a degree semantics compared to languages with a degree semantics are important to see how gradability and comparison are expressed and understood. To determine whether the conclusions about the interpretation of relative and absolute gradable adjectives can be generalized to other members of these classes, adjectives from other dimensions, e.g., *lang-kurz* ('long'-'short') and *nass-trocken* ('wet'-'dry'), should be investigated. Further studies are necessary to compare the interpretation of gradable adjectives with regard to existing and non-existing objects, and with regard to three-dimensional objects and objects displayed on picture cards. The comparison of existing and non-existing objects could reveal whether the interpretation of gradable adjectives differs depending on children's knowledge about real-world objects and their properties in contrast to novel objects for which participants do not know anything about their properties. Gradable properties of three-dimensional objects may be easier to perceive than gradable properties of objects displayed on picture cards. The presentation of physical objects may particularly improve the performance of 3-year-olds.

The present forced picture-choice task turned out to be a suitable experimental design for children as young as age 3 and could be extended to the aspects mentioned above. In contrast to most previous acquisition studies on the interpretation of gradable adjectives, changes in the comparison class were expressed by using different nouns (basic-level and superordinate-level nouns). The effects of these changes were analyzed within subjects rather than between subjects as in previous studies. The current design has several advantages as explained in Chapter 5. For instance, it could be assessed whether changes in the noun affect the determination of the comparison class for identical visual displays. However, this design may have also led to unexpected response patterns that are possi-

bly related to the superordinate-level noun *Spielsachen* 'toys'. Take for instance the test prompt *Gib mir bitte die großen Spielsachen* ('Please give me the big toys'). Some particpants calculated two separate standards of comparison for each basic-level category, e.g., one for soccer balls and one for water balloons, instead of one joint standard for the whole category of toys. Another unexpected response was to not have different standards for e.g., *big water balloons* and *big toys* although the comparison class differed. Three modifications of the experimental design may eliminate these unexpected responses. First, instead of superordinate- and basic-level nouns the test prompts could include basic-level and subordinate-level nouns. This means, the participants would see soccer balls and tennis balls and are asked once to give the big balls and another time to give the big tennis balls. This modification may reduce the calculation of two separate standards because it may be more natural two subsume tennis balls and soccer balls under the basic-level noun *Bälle* ('balls') than to subsume water balloons and soccer balls under the superordinate-level noun *Spielsachen* ('toys'). However, this modification has the disadvantage that for instance tennis balls are less likely to differ in size than water balloons are – something that should be the case given that they are displayed in different sizes in the visual context. As a second modification of the experimental design the test prompts could always include the same noun, e.g., *Spielsachen* ('toys'), in combination with different visual contexts. This means participants are for instance asked to give the big toys while seeing once only water balloons and another time water balloons and soccer balls. This modification may prompt more changes of the standard of comparison than in the present design in which the visual context did not change. However, with the modified design it is not possible to figure out whether participants use the noun to determine the comparison class. The third possibility would be to use a between-subjects rather than a within-subjects design (see Barner & Snedeker, 2008) with the result that the interpretation of one item cannot influence the interpretation of another item. On the other hand, a between-subjects design cannot reveal the direct effect of changes in the comparison class on an individual's interpretation of gradable adjectives.

Future studies should explore the role of the input for the acquisition of adjectives, they should consider a greater range of semantic properties, and they should investigate the relation between semantic properties of adjectives and other domains. Regarding the role of the input, the findings of this thesis can serve as a starting point for a more extensive analysis of the input. Moreover, it is a crucial question which cues the input provides to enable the child to acquire the semantic properties of adjectives. The *Semantic Complexity Hierarchy* developed in this thesis can serve as a starting point to derive hypotheses about the acquisition of other adjective classes such as gradable evaluative adjectives or non-subsective/non-intersective adjectives. Further research on the semantics of adjectives is necessary to confirm the classification of adjectives as proposed in the *Semantic Complexity Hierarchy* and to be able to extend it to other adjectival modifiers such as participles or ordinal numbers. This thesis also addressed the relation between semantics and other domains. The findings of the spontaneous speech data analysis point to a relation between

semantic properties and the syntactic position of adjectives (see also Blackwell, 2001, for a relation between notional properties and syntactic position). It is unclear why some adjective classes were preferably produced in attributive position whereas other adjective classes were preferably produced in predicative position. Studies with a more detailed syntactic analysis that take also semantic properties of adjectives into account are required. Moreover, a relationship between semantic and notional properties of adjectives is likely. The coding of the adjectives produced by Leo according to both semantic and conceptual criteria is a first step in understanding this relation, which is also accounted for in the proposed acquisition path. Further research is needed to explore the relation between the acquisition of meaning and the development of concepts.

To summarize, this thesis combined theoretical and empirical work in the area of the semantics of adjectives. The findings reported show that formal semantic properties of adjectives are mirrored in psycholinguistic data from language acquisition. These findings may launch further theoretical and acquisition research on adjectives and other modification structures. In the long run, it would be desirable to explore how other populations than monolingual typically-developing German-speaking children acquire the semantics and syntax of adjectives, in which domains additional support makes sense, and how adjectives should be framed in the input to be beneficial for the language learner.

References

Alexiadou, A., Haegeman, L., & Stavrou, M. (2007). *Noun phrase in the generative perspective.* Berlin: De Gruyter.

Aparicio, H., Kennedy, C., & Xiang, M. (2018). Perceived Informativity and Referential Effects of Contrast in Adjectivally Modified NPs. In E. Castroviejo, L. McNally, & G. Sassoon (Eds.), *The Semantics of Gradability, Vagueness, and Scale Structure. Experimental perspectives* (p. 199-220). Berlin: Springer.

Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret "tall" and "short" based on the size distributions of novel noun referents. *Child Development*, *79*(3), 594-608.

Bartlett, E. (1976). Sizing things up: the acquisiton of the meaning of dimensional adjectives. *Journal of Child Language*, *3*(2), 205-219.

Beck, S., Krasikova, S., Fleischer, D., Gergel, R., Savelsberg, C., Vanderelst, J., & Villalta, E. (2009). Crosslinguistic variation in comparison constructions. In O. Percus, H.-M. Gärtner, & S. Beck (Eds.), *Linguistic variation yearbook 9* (p. 1-66). Amsterdam: John Benjamins.

Becker, M. (2000). *The Development of the Copula in Child English: The Lightness of 'Be'* (Unpublished doctoral dissertation). University of Pennsylvania.

Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes*, *21* , 2-24.

Belke, E. (2001). *On the time course of naming multidimensional objects in a referential communication task* (Unpublished doctoral dissertation). University of Bielefeld.

Bhatt, R. (1999). *Covert modality in non-finite contexts* (Unpublished doctoral dissertation). University of Pennsylvania.

Bierwisch, M. (1987). Dimensionsadjektive als strukturierender Ausschnitt des Sprachverhaltens. In M. Bierwisch & E. Lang (Eds.), *Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven* (p. 91-286). Berlin: Akademie-Verlag.

Bierwisch, M. (1988). On the grammar of local prepositions. In M. Bierwisch, W. Motsch, & I. Zimmermann (Eds.), *Syntax, Semantik und Lexikon* (p. 1-65). Berlin: Akademie-Verlag.

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation* (p. 71-261). Berlin: Springer.

Bittner, D. (1998). Entfaltung grammatischer Relationen im NP-Erwerb: Referenz. *Folia Linguistica*, *XXXI*, 255-283.

Bittner, D. (2002). Emergence of grammatical complexity and markedness in the acquisition of Verb and Noun Phrases in German. In K. Dziubalka-Kolaczyk & J. Weckwerth (Eds.), *Future challenges for natural linguistics* (p. 25-56). Wien: Lincom.

Blackwell, A. (2001). On the Acquisition of the Syntax of English Adjectives. In A. Okrent & J. Boyle (Eds.), *CLS 36, Volume 2: The Panels.* Chicago: University of Chicago

Press.

Blackwell, A. (2005). Acquiring the English adjective lexicon: relationships with input properties and adjectival semantic typology. *Journal of Child Language*, *32*(3), 535-562.

Blackwell, A., & Olson, C. (2008). The nature of nominals modified by adjectives in the input. In T. Marinis, A. Papangeli, & V. Stojanovik (Eds.), *Proceedings of the Child Language Seminar 2007 - 30th Anniversary* (p. 38-48). Retrieved from http://www.reading.ac.uk/web/FILES/cls/CLS_Blackwell,Olson.pdf.

Bochnak, R. (2015). The Degree Semantics Parameter and cross-linguistic variation. *Semantics & Pragmatics*, *8*, 1-48.

Bolinger, D. (1967). Adjectives in English: Attribution and Predication. *Lingua*, *18*, 1-34.

Booij, E., & Sassoon, G. W. (2014). Big differences: The standard for 'big' as used by adults and children. In N. Melnik (Ed.), *Proceedings of the Israel Association of Theroretical Linguistics 29* (p. 1-14). MIT Working Papers in Linguistics 72.

Bowers, J. (1993). The syntax of predication. *Linguistic Inquiry*, *24*(4), 591-656.

Bowler, M. (2016). The status of degrees in Warlpiri. In M. Grubic & A. Mucha (Eds.), *Proceedings of 'The Semantics of African, Asian and Austronesian Languages' 2* (p. 1-17). Potsdam: Universitätsverlag Potsdam.

Britton, M., Ipesa-Balogun, H., Jackson, E., Kotfila, J., Lu, K., Prabhakar, V., ... de Villiers, P. (2017). *Acquisition of the distinction between absolute and relative gradable adjectives: Four-year-olds know emotion adjectives are relative* (Poster presented at the Society for Research in Child Development Biennial Meeting). Austin, Texas.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language.* New York, London, Sidney, Toronto: Wiley.

Bußmann, H. (2008). *Lexikon der Sprachwissenschaft* (4th ed.). Stuttgart: Kröner.

Bylinina, L. (2013). *The grammar of standards - judge-dependence, purpose-relativity, and comparison classes in degree constructions* (Unpublished doctoral dissertation). Utrecht University.

Bylinina, L., Ivlieva, N., Podobryaev, A., & Sudo, Y. (2015). An in situ semantics for ordinals. In T. Bui & D. Özyıldız (Eds.), *NELS 45: Proceedings of the Forty-Fifth Annual Meeting of the North East Linguistic Society: Volume 1* (p. 135-145). CreateSpace Independent Publishing Platform.

Cabredo Hofherr, P. (2010). Adjectives - an introduction. In P. C. Hofherr & O. Matushansky (Eds.), *Adjectives – Formal analyses in syntax and semantics* (p. 1-26). Amsterdam: John Benjamins.

Callanan, M. A., & Markman, E. M. (1982). Principles of organization in young children's natural language hierarchies. *Child Development*, *53*(4), 1093-1101.

Castroviejo, E., McNally, L., & Sassoon, G. W. (2018). Gradability, Vagueness, and Scale Structure: From the Armchair to the Lab. In E. Castroviejo, L. McNally, & G. W. Sassoon (Eds.), *The Semantics of Gradability, Vagueness, and Scale Structure:*

Experimental Perspectives (p. 1-24). Berlin: Springer.

Cinque, G. (2010). *The syntax of adjectives. A comparative study.* Cambridge, MA: MIT Press.

Clahsen, H., Eisenbeiss, S., & Vainikka, A. (1994). The seeds of structure. A syntactic analysis of the acquisition of case marking. In T. Hoekstra & B. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar* (p. 85-118). Amsterdam: John Benjamins.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (p. 65-110). New York, London: Academic Press.

Clark, E. V. (1993). *The lexicon in acquisition.* Cambridge, MA: Cambridge University Press.

Clark, E. V. (2004). How language acquisition builds on cognitive development. *TRENDS in Cognitive Sciences*, *8*(10), 472-478.

Clark, H. H. (1970). The primitive nature of children's relational concepts. In J. Hayes (Ed.), *Cognition and the development of language* (p. 269-278). New York: Wiley.

Cresswell, M. (1976). The semantics of degree. In B. Partee (Ed.), *Montague grammar* (p. 261-292). New York: Academic Press.

Culicover, P. W. (2013). *Grammar & complexity.* Oxford: Oxford University Press.

Culicover, P. W. (2014). Constructions, complexity, and word order variation. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity* (p. 148-178). Oxford: Oxford University Press.

Davidson, D. (1967). The logical form of action sentences. In N. Resher (Ed.), *The logic of decision and action* (p. 81-120). Pittsburgh: University of Pittsburgh Press.

Davies, C., Lingwood, J., & Arunachalam, S. (2019). Adjective forms and functions in British English child-directed speech. *Journal of Child Language*, 1-27.

Deal, A. R., & Hohaus, V. (2019). Vague predicates, crisp judgements. In M. Espinal, E. Castroviejo, M. Leonetti, L. McNally, & C. Real-Puigdollers (Eds.), *Proceedings of Sinn und Bedeutung 23* (p. 347-364). Retrieved from https://semanticsarchive.net/Archive/Tg3ZGI2M/Proceedings23.html.

Demonte, V. (2011). Adjectives. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 2, p. 1314-1340). Berlin: De Gruyter.

Demuth, K. (1998). Collecting spontaneous production data. In D. McDaniel, C. McKee, & H. Cairns (Eds.), *Methods for assessing children's syntax* (p. 3-22). Cambridge, MA: MIT Press.

den Dikken, M. (2006). *Relators and Linkers. The syntax of predication, predicate inversion, and copulas.* Cambridge, MA: MIT Press.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*, 159-184.

Dixon, R. M. W. (1982). *'Where have all the adjectives gone?' and other essays in semantics and syntax.* Berlin: Mouton.

Dixon, R. M. W. (2004). Adjective classes in typological perspective. In R. M. W. Dixon & A. Aikhenvald (Eds.), *Adjective classes. a cross-linguistic typology.* Oxford: Oxford University Press.

Dromi, E. (1987). *Early lexical development.* Cambridge: Cambridge University Press.

Dromi, E. (1999). Early lexical development. In M. Barrett (Ed.), *The development of language* (p. 99-131). Hove: Psychology Press.

Ebeling, K., & Gelman, S. (1988). Coordination of size standards by young children. *Child Development*, *59*(4), 888-869.

Ebeling, K., & Gelman, S. (1994). Children's use of context in interpreting "big" and "little". *Child Development*, *65*(4), 1178-1192.

Ehri, L. (1976). Comprehension and production of adjectives and seriation. *Journal of Child Language*, *3*, 369-384.

Eilers, R., Oller, D., & Ellington, J. (1974). The acquisition of word-meaning for dimensional adjectives: the long and short of it. *Journal of Child Language*, *1*(2), 195-204.

Eisenbeiss, S. (2000). The acquisition of the Determiner Phrase in German child language. In M.-A. Friedemann & L. Rizzi (Eds.), *The Acquisition of Syntax: Studies in Comparative Developmental Linguistics* (p. 26-62). Harlow: Longman.

Fernald, A., Thorpe, K., & Marchmann, V. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology*, *60*, 190-217.

Field, A. (2009). *Discovering statistics using SPSS.* London: SAGE.

Foppolo, F., & Panzeri, F. (2013). Do children know when their room counts as clean? In S. Kan, C. M. Cantwell, & R. Staubs (Eds.), *NELS 40: Proceedings of the Fortieth Annual Meeting of the North East Linguistic Society: Volume 1* (p. 205-218). Amherst, MA: GLSA.

Frank, R. (1998). Structural complexity and the time course of grammatical development. *Cognition*, *66*, 249-301.

Frawley, W. (1992). *Linguistic semantics.* Hillsdale, NJ: Erlbaum.

Frazier, L., Clifton, C., & Stolterfoth, B. (2008). Scale structure: Processing minimum standard and maximum standard scalar adjectives. *Cognition*, *106*, 299-324.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, *31*, 661-681.

Gao, H., Zelazo, P., Sharpe, D., & Mashari, A. (2014). Beyond early linguistic competence: Development of children's ability to interpret adjectives flexibly. *Cognitive Development*, *32*, 86-102.

Gelman, S., & O'Reilly, A. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development*, *59*(4), 876-887.

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, *89*, 27-71.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development, Vol. 2: Language thought and culture* (p. 301-334). Hillsdale, NJ: Erlbaum.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3-55.

Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5). Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen.* Göttingen: Hogrefe.

Hamann, C. (1991). Adjectives. In A. von Stechow & D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung* (p. 657-673). Berlin, New York: De Gruyter.

Hansen, N., & Chemla, E. (2017). Color adjectives, standards, and thresholds: an experimental investigation. *Linguistics and Philosophy*, *40*(3), 239-278.

Hawkins, J. (1987). Implicational universals as predictors of language acquisition. *Linguistics*, *25*, 453-473.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar.* Malden, MA: Blackwell.

Heller, D., & Chambers, C. (2014). Would a blue kite by any other name be just as blue? Effects of descriptive choices on subsequent referential behavior. *Journal of Memory and Language*, *70*, 53-67.

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning. In R. M. Golinkoff et al. (Eds.), *Becoming a word learner: A debate on lexical acquisition* (p. 136-164). Oxford: Oxford University Press.

Hohaus, V., Tiemann, S., & Beck, S. (2014). Acquisition of comparison constructions. *Language Acquisition*, *21*(3), 215-249.

Huang, Y., & Snedeker, J. (2013). The use of referential context in children's on-line interpretation of scalar adjectives. *Developmental Psychology*, *49*, 1090-1102.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*(6), 597-626.

Jackendoff, R. (1983). *Semantics and Cognition.* Cambridge, MA: MIT Press.

Jackendoff, R. (1990). *Semantic structure.* Cambridge, MA: MIT Press.

Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics of natural language* (p. 123-155). Cambridge: Cambridge University Press.

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, *57*, 129-191.

Kamp, H., & Reyle, U. (1993). *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Dordrecht: Springer.

Kanemura, H., Aihara, M., Aoki, S., Araki, T., & Nakazawa, S. (2003). Development of the prefrontal lobe in infants and children: a threedimensional magnetic resonance volumetric study. *Brain & Development*, *25*, 195-199.

Kauschke, C. (2000). *Der Erwerb des frühkindlichen Lexikons: Eine empirische Studie zur Entwicklung des Wortschatzes im Deutschen.* Tübingen: Narr.

Kauschke, C. (2007). *Erwerb und Verarbeitung von Nomen und Verben.* Tübingen: Niemeyer.

Kennedy, C. (2006). Comparatives, Semantics of. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (p. 690-694). Amsterdam: Elsevier.

Kennedy, C. (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1-45.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*(2), 345-381.

Kennedy, C., & McNally, L. (2010). Color, context and compositionality. *Synthese*, *174*, 79-98.

Klatzky, R. L., Clark, E. V., & Macken, M. (1973). Asymmetries in the acquisition of polar adjectives: Linguistic or conceptual? *Journal of Experimental Child Psychology*, *16*, 32-46.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*, 1-45.

Klibanoff, R., & Waxman, S. (2000). Basic level object categories support the acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development*, *71*(3), 649-659.

Kratzer, A. (1995). Individual-level and stage-level predicates. In G. N. Carlson & F. J. Pelletier (Eds.), *The generic book* (p. 125-175). Chicago: Chicago University Press.

Laenzlinger, C. (2005). French adjective ordering: perspectives on DP-internal movement types. *Lingua*, *115*, 645-689.

Lakusta, L., & DiFabrizio, S. (2016). And, the winner is... a visual preference for end points over starting points in infants' motion event representations. *Infancy*, 1-21. doi: 10.1111/infa.12153

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*, 1-33.

Lapata, M., & Keller, F. (2005). Web-based models for natural language processing. *ACM Transactions on Speech and Language Processing*, *2*, 1-30.

Larson, R. (1999). *Semantics of adjectival modification* (Tech. Rep.). (Lecture Notes, LOT Winter School, Amsterdam)

Lasersohn, P. (2005). Context-dependence, disagreement and predicates of personal taste. *Linguistics and Philosophy*, *28*, 643-686.

Liao, S., & Meskin, A. (2015). Aesthetic adjectives: Experimental semantics and context-sensitivity. *Philosophy and Phenomenological Research*, 1-28.

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics*, *1*, 12.1-12.21.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed., Vol. 2: The Database). Mahwah, NJ: Lawrence Erlbaum Associates.

Maienborn, C. (2003). *Die Logische Form von Kopula-Sätzen.* Berlin: Akademie-Verlag.

Mandler, J. M. (2012). On the spatial foundations of the conceptual system and its enrichment. *Cognitive Science*, *36*, 421-451.

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57-77.

Matthewson, L. (2014). The measurement of semantic complexity: how to get by if your language lacks generalized quantifiers. In F. J.Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity* (p. 241-263). Oxford: Oxford University Press.

McNally, L. (2011). The relative role of property type and scale structure in explaining the behavior of gradable adjectives. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *ViC 2009 (Papers from the ESSLLI 2009 Workshop on Vagueness in Communcation)* (p. 151-168). Berlin: Springer.

McNally, L. (2016). Modification. In M. Aloni & P. Dekker (Eds.), *Cambridge Handbook of Semantics.* Cambridge: Cambridge University Press.

McNally, L., & Stojanovic, I. (2017). Aesthetic adjectives. In J. Young (Ed.), *Semantics of aesthetic judgment.* Oxford: Oxford University Press.

Menn, L., & Duffield, C. (2014). Looking for a 'gold standard' to measure language complexity: what psycholinguistics and neurolinguistics can (and cannot) offer to formal linguistics. In F. Newmeyer & L. Preston (Eds.), *Measuring grammatical complexity.* Oxford University Press.

Menyuk, P., Liebergott, J. W., & Schultz, M. C. (1995). *Early language development in full term and premature infants.* Hillsdale, NJ: Erlbaum.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology* (Vol. 32, p. 89-115).

Mintz, T. (2005). Linguistic and conceptual influences on adjective acquisition in 24- and 36-month-olds. *Developmental Psychology*, *41*(1), 17-29.

Mintz, T., & Gleitman, L. (2002). Adjectives do really modify nouns: the incremental and restricted nature of early adjective acquisition. *Cognition*, *84*, 267-293.

Montague, R. (1970). English as a formal language. In B. Visentini (Ed.), *Linguaggi nella societa e nella tecnica* (p. 189-224). Milan: Edizioni di Comunita.

Morzycki, M. (2012). Adjectival extremeness: degree modification and contextually restricted scales. *Natural Language and Linguistic Theory*, *30*, 567-609.

Morzycki, M. (2014). *Where does nominal gradability come from?* (Paper presented at University of Massachusetts)

Morzycki, M. (2016). *Modification.* Cambridge: Cambridge University Press.

Morzycki, M. (to appear). Structure and ontology in nonlocal readings of adjectives. In T. Gamerschlag & S. Löbner (Eds.), *Cognitive structures: Linguistic, philosophical,*

and psychological perspectives. Dordrecht: Springer.

Moscati, V., & Rizzi, L. (2014). Agreement configurations in language development: A movement-based complexity metric. *Lingua*, *140*, 67-82.

Müller, A. (2012). *Wie interpretieren Kinder 'nur'? Experimentelle Untersuchungen zum Erwerb von Informationsstruktur.* Universitätsverlag Potsdam.

Müller, A., Schulz, P., & Tracy, R. (2018). Spracherwerb. In C. Titz, S. Geyer, A. Ropeter, H. Wagner, S. Weber, & M. Hasselhorn (Eds.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln* (p. 53-68). Stuttgart: Kohlhammer.

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology*, *15*(4), 572-586.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*, 1-135.

Nelson, K. (1976). Some attributes of adjectives used by young children. *Cognition*, *4*, 13-30.

Nelson, K., & Benedict, H. (1974). The Comprehension of Relative, Absolute and Contrastive Adjectives by Young Children. *Journal of Psycholinguistic Research*, *3*(4), 333-342.

Ninio, A. (2004). Young children's difficulty with adjectives modifying nouns. *Journal of Child Language*, *31*, 255-285.

Panzeri, F., & Foppolo, F. (2012). Can children tell us something about the semantics of adjectives? In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, & M. Westera (Eds.), *Logic, Language and Meaning. Lecture Notes in Computer Science* (p. 170-179). Berlin, Heidelberg: Springer.

Partee, B. (1995). Lexical Semantics and Compositionality. In D. Osherson, L. Gleitman, & M. Liberman (Eds.), *An invitation to Cognitive Science: Part 1 Language* (2nd ed., p. 311-360). Cambridge, MA: MIT Press.

Partee, B. (2009). Formal semantics, lexical semantics, and compositionality: The puzzle of privative adjectives. *Philologia*, *7*, 11-23.

Partee, B., & Hendriks, H. (1997). Montague grammar. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (p. 5-91). Amsterdam: Elsevier.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*, 89-110.

Piaget, J., & Inhelder, B. (1973). *Die Entwicklung der elementaren logischen Strukturen.* Düsseldorf: Pädag. Verl. Schwann.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Rapp, I. (2015). On the Temporal Interpretation of Present Participles in German. *Journal of Semantics*, *32*(3), 477–523.

Rips, L., & Turnbull, W. (1980). How big is big? Relative and absolute properties in memory. *Cognition*, *8*, 145-174.

Roget, P. M. (1965). *The St. Martin's Roget's Thesaurus of English Words and Phrases.* New York: St. Martin's Press.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

Rothstein, S. (1999). Fine-grained structure in the eventuality domain: the semantics of predicate adjective phrases and 'be'. *Natural Language Semantics*, *7*, 347-420.

Rothweiler, M. (2007). Der Aufbau des kindlichen Lexikons. In J. Meibauer et al. (Eds.), *Einführung in die germanistische Linguistik* (p. 266-272). Stuttgart: J. B. Metzler.

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, *12*(3), 259-288.

Ryalls, B., & Smith, L. (2000). Adult's acquisition of novel dimension words: Creating a semantic congruity effect. *The Journal of General Psychology*, *127*(3), 279-326.

Sandhofer, C., & Smith, L. (2007). Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language Learning and Development*, *3*(3), 233-267.

Sandhofer, C., & Smith, L. B. (2001). Why children learn color and size words so differently: Evidence from adults' learning of artificial terms. *Journal of Experimental Psychology: General*, *130*(4), 600-620.

Sapir, E. (1944). Grading, a study in semantics. *Philosophy of Science*, *11*(2), 93-116.

Sassoon, G. W. (2013). A typology of multidimensional adjectives. *Journal of Semantics*, *30*, 335-380.

Sassoon, G. W. (2017). Comparisons of nominal degrees. *Language*, *93*(1), 153-188.

Sassoon, G. W. (in preparation). Multidimensionality in the grammar of gradability.

Saylor, M. (2000). Time-stability and adjective use by child and adult English speakers. *First Language*, *20*, 91-120.

Schulz, P. (2018). Zur Semantik von Verben im Spracherwerb. In S. Engelberg, H. Lobin, K. Steyer, & S. Wolfer (Eds.), *Wortschätze. Dynamik – Muster – Komplexität. (= Institut für Deutsche Sprache, Jahrbuch 2017)* (p. 133-151). Berlin, New York: De Gruyter.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through context representation. *Cognition*, *71*, 109-147.

Sera, M., & Smith, L. B. (1987). Big and little: 'nominal' and relative uses. *Cognitive Development*, *2*, 89-111.

Siegel, L. (1977). The basis of the comprehension and production of relational terminology. *Journal of Experimental Child Psychology*, *24*, 40-52.

Smith, L. B., Cooney, N., & McCord, C. (1986). What is high? The development of reference points for high and low. *Child Development*, *57*, 583-602.

Smith, L. B., Gasser, M., & Sandhofer, C. M. (1997). Learning to talk about the properties of objects: A network model of the development of dimensions. In R. L. Goldstone, D. L. Medin, & P. G. Schyns (Eds.), *Perceptual learning. The psychology of learning*

and motivation (p. 219-255). San Diego, CA: Academic Press.

Snyder, W. (2007). *Child language*. Oxford: Oxford University Press.

Solt, S. (2011). Notes on the comparison class. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *ViC 2009 (Papers from the ESSLLI 2009 Workshop on Vagueness in Communcation)* (p. 189-206). Berlin: Springer.

Solt, S. (2016). Ordering subjectivity and the absolute/relative distinction. In N. Bade, P. Berezovskaya, & A. Schöller (Eds.), *Proceedings of Sinn und Bedeutung 20* (p. 676-693). Retrieved from https://semanticsarchive.net/Archive/GRmOGQ4N/SUB20html4.html.

Solt, S. (2018). Multidimensionality, subjectivity and scales: Experimental evidence. In E. Castroviejo, L. McNally, & G. W. Sassoon (Eds.), *The Semantics of Gradability, Vagueness and Scale Structure: Experimental Perspectives* (p. 59-92). Berlin: Springer.

Solt, S., & Gotzner, N. (2012). Experimenting with degree. In A. Chereches (Ed.), *Proceedings of the 22nd Semantics and Linguistic Theory Conference (SALT 22)* (p. 166-187). Linguistic Society of America.

Sproat, R., & Shih, C. (1991). The cross-linguistic distributions of adjective ordering restrictions. In C. Georgopoulos & R. Ishihara (Eds.), *Interdisciplinary Approaches to Language* (p. 565-593). Dordrecht: Springer.

Stowell, T. (1978). What was there before 'there' was there? In D. Farkas, W. M. Jacobsen, & K. W. Todrys (Eds.), *Papers from the Fourteenth Regional Meeting of the Chicago Linguistics Society* (p. 457-471). Chicago: Chicago Linguistic Society.

Stromswold, K. (1998). Analyzing children's spontaneous speech. In D. McDaniel, C. McKee, & H. Cairns (Eds.), *Methods for assessing children's syntax* (p. 23-53). Cambridge, MA: MIT Press.

Syrett, K. (2007). *Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives* (Unpublished doctoral dissertation). Northwestern University.

Syrett, K., Bradley, E., Kennedy, C., & Lidz, J. (2006). Shifting Standards: Children's Understanding of Gradable Adjectives. In K. U. Deen, J. Nomura, B. Schulz, & B. Schwartz (Eds.), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition - North America (GALANA)* (Vol. 4, p. 353-364). Cambridge, MA: UConn Occasional Papers in Linguistic.

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and Context in Children's Understanding of Gradable Adjectives. *Journal of Semantics*, *27*, 1-35.

Taylor, M., & Gelman, S. (1988). Adjective and Nouns: Children's Strategies for Learning New Words. *Child Development*, *59*(2), 411-419.

Thorpe, K., Baumgartner, H., & Fernald, A. (2006). Children's Developing Ability to Interpret Adjective-Noun Combinations. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *BUCLD 30: Proceedings of the 30th Annual Boston University Conference on Language Development* (Vol. 2, p. 631-642). Somerville, MA: Cascadilla Press.

Toledo, A., & Sassoon, G. W. (2011). Absolute vs. relative adjectives – variance within vs. between individuals. In N. Ashton, A. Chereches, & D. Lutz (Eds.), *Proceedings of the 21st Semantics and Linguistic Theory Conference (SALT 21).* Linguistic Society of America.

Trabandt, C. (2017). *On the acquisition of restrictive and appositive relative clauses* (Doctoral dissertation). Goethe-University Frankfurt, Retrieved from http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/45273.

Tribushinina, E. (2013). Adjective Semantics, World Knowledge and Visual Context: Comprehension of Size Terms by 2- to 7-Year-Old Dutch-Speaking Children. *Journal of Psycholinguistic Research*, *42*, 205-225.

Tribushinina, E. (2017). Do degree adverbs guide adjective learning cross-linguistically? A comparison of Dutch and Russian. *Linguistics*, *55*(4), 899-933.

Tribushinina, E., & Gillis, S. (2012). The acquisition of scalar structures: Production of adjectives and degree markers by Dutch-speaking children and their caregivers. *Linguistics*, *50*(2), 241 – 268.

Tribushinina, E., Gillis, S., & De Maeyer, S. (2013). Infrequent word classes in the speech of two- to seven-year-old children with cochlear implants and their normally hearing peers: A longitudinal study of adjective use. *Journal of Pediatric Otorhinolaryngology*, *77*, 356-361.

Tribushinina, E., van den Bergh, H., Kilani-Schoch, M., Aksu-Koc, A., Dabasinskiene, I., Hrzica, G., ... Dressler, W. (2013). The role of explicit contrast in adjective acquisition: A crosslinguistic longitudinal study of adjective production in spontaneous child speech and parental input. *First Language*, *33*(6), 594–616.

Tribushinina, E., van den Bergh, H., Ravid, D., Aksu-Koç, A., Kilani-Schoch, M., Korecky-Kröll, K., ... Gillis, S. (2014). Development of adjective frequencies across semantic classes. A growth curve analysis of child speech and child-directed speech. *Language, Interaction and Acquisition*, *5*, 185-226.

Umbach, C. (2016). Evaluative propositions. In C. Meier & J. van Wijnbergen-Huitink (Eds.), *Subjective meaning* (p. 127-168). Berlin: De Gruyter.

Ünal, E., & Papafragou, A. (2016). Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, *89*, 179-199.

Van Geenhoven, V. (2006). Acquisition and interpretations: A brief introduction. In V. Van Geenhoven (Ed.), *Semantics in acquisition* (p. 1-18). Dordrecht: Springer.

von Fintel, K., & Heim, I. (1999). *More on lousy teachers and beautiful dancers.* (Unpublished manuscript.)

von Stechow, A. (1984). Comparing semantic theories of comparison. *Journal of Semantics*, *3*, 1-77.

Wagner, K., Storage, D., Dobkins, K., & Barner, D. (2013). *Children's acquisition of subsective and intersective adjectives* (Poster presented at BUCLD 38). George Sherman Union; Boston University.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes San-

tos, M., ... Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*(2), 119-125.

Waxman, S., & Booth, A. (2001). Seeing Pink Elephants: Fourteen-Month-Olds' Interpretations of Novel Nouns and Adjectives. *Cognitive Psychology*, *43*, 217-242.

Waxman, S., Shipley, E., & Shepperson, B. (1991). Establishing new subcategories: The role of category labels and existing knowledge. *Child Development*, *62*, 127-138.

Weicker, M., & Schulz, P. (2018). Is 'clean' the Same as 'not dirty'? On the Understanding of Absolute Gradable Adjectives. In A. B. Bertolini & M. J. Kaplan (Eds.), *BUCLD 42: Proceedings of the 42nd Annual Boston University Conference on Language Development* (p. 790-802). Somerville, MA: Cascadilla Press.

Weicker, M., & Schulz, P. (2019). Red train, big train, broken train - semantic aspects of adjectives in child language. In M. Rispoli & T. Ionin (Eds.), *Three Streams of Generative Language Acquisition Research: Selected Papers from the 7th Meeting of Generative Approaches to Language Acquisition – North America, University of Illinois at Urbana-Champaign.* Amsterdam: John Benjamins.

Wiese, H. (2003). *Numbers, language, and the human mind.* Cambridge: Cambridge University Press.

Ziegler, J., & Pylkkänen, L. (2016). Scalar adjectives and the temporal unfolding of semantic composition: An MEG investigation. *Neuropsychologia*, *89*, 161-171.

Appendix A

Study 1

A.1 Adjectives in the Leo-corpus

The following table lists all adjectives produced by Leo between 2;00 and 2;11 years.

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
beige	1	COLOR
silbern	1	COLOR
blau	1	COLOR
grün	1	COLOR
rosa	1	COLOR
rot	1	COLOR
lila	1	COLOR
gelb	1	COLOR
braun	1	COLOR
weiß	1	COLOR
schwarz	1	COLOR
bunt	1	COLOR
orange	1	COLOR
gold	1	COLOR
grau	1	COLOR
türkis	1	COLOR
wahr	1	CONFORMITY
original	1	CONFORMITY
richtig	1	CONFORMITY
falsch	1	CONFORMITY
echt	1	CONFORMITY
hessisch	1	NATIONALITY

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
englisch	1	NATIONALITY
russisch	1	NATIONALITY
französisch	1	NATIONALITY
chinesisch	1	NATIONALITY
deutsch	1	NATIONALITY
amerikanisch	1	NATIONALITY
österreichisch	1	NATIONALITY
sächsisch	1	NATIONALITY
afrikanisch	1	NATIONALITY
ägyptisch	1	NATIONALITY
neuneckig	1	PHYSICAL PROPERTY
eckig	1	PHYSICAL PROPERTY
dreieckig	1	PHYSICAL PROPERTY
tot	1	PHYSICAL STATE
stumm	1	PHYSICAL STATE
blind	1	PHYSICAL STATE
schief	2a	PHYSICAL PROPERTY
kurvig	2a	PHYSICAL PROPERTY
zottelig	2a	PHYSICAL PROPERTY
strubbelig	2a	PHYSICAL PROPERTY
flauschig	2a	PHYSICAL PROPERTY
locker	2a	PHYSICAL PROPERTY
schlaff	2a	PHYSICAL PROPERTY
bitter	2a	PHYSICAL PROPERTY
dreckig	2a	PHYSICAL PROPERTY
sauber	2a	PHYSICAL PROPERTY
wach	2a	PHYSICAL PROPERTY
nass	2a	PHYSICAL PROPERTY
trocken	2a	PHYSICAL PROPERTY
leer	2a	PHYSICAL PROPERTY
offen	2a	PHYSICAL PROPERTY
kaputt	2a	PHYSICAL PROPERTY
platt	2a	PHYSICAL PROPERTY
voll	2a	PHYSICAL PROPERTY
heil	2a	PHYSICAL PROPERTY
wackelig	2a	PHYSICAL PROPERTY
salzig	2a	PHYSICAL PROPERTY

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
rund	2a	PHYSICAL PROPERTY
glatt	2a	PHYSICAL PROPERTY
fest	2a	PHYSICAL PROPERTY
flach	2a	PHYSICAL PROPERTY
schmutzig	2a	PHYSICAL PROPERTY
gestreift	2a	PHYSICAL PROPERTY
nackt	2a	PHYSICAL PROPERTY
süß	2a	PHYSICAL PROPERTY
scharf	2a	PHYSICAL PROPERTY
krumm	2a	PHYSICAL PROPERTY
gerade	2a	PHYSICAL PROPERTY
spitz	2a	PHYSICAL PROPERTY
stabil	2a	PHYSICAL PROPERTY
rutschig	2a	PHYSICAL PROPERTY
kahl	2a	PHYSICAL PROPERTY
stachelig	2a	PHYSICAL PROPERTY
durchsichtig	2a	PHYSICAL PROPERTY
gepunktet	2a	PHYSICAL PROPERTY
kariert	2a	PHYSICAL PROPERTY
feucht	2a	PHYSICAL PROPERTY
hungrig	2a	PHYSICAL STATE
müde	2a	PHYSICAL STATE
satt	2a	PHYSICAL STATE
heiser	2a	PHYSICAL STATE
blass	2a	PHYSICAL STATE
durstig	2a	PHYSICAL STATE
schwindelig	2a	PHYSICAL STATE
kitzelig	2a	PHYSICAL STATE
wasserscheu	2a	BEHAVIOR
teuer	2b	n.d.
niedrig	2b	DIMENSION
leicht	2b	DIMENSION
groß	2b	DIMENSION
klein	2b	DIMENSION
dick	2b	DIMENSION
hoch	2b	DIMENSION
kurz	2b	DIMENSION

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
schwer	2b	DIMENSION
eng	2b	DIMENSION
tief	2b	DIMENSION
lang	2b	DIMENSION
schmal	2b	DIMENSION
weit	2b	DIMENSION
breit	2b	DIMENSION
fern	2b	DIMENSION
nah	2b	DIMENSION
dünn	2b	DIMENSION
flink	2b	PHYSICAL PROPERTY
stark	2b	PHYSICAL PROPERTY
schwach	2b	PHYSICAL PROPERTY
schnell	2b	PHYSICAL PROPERTY
weich	2b	PHYSICAL PROPERTY
hart	2b	PHYSICAL PROPERTY
dunkel	2b	PHYSICAL PROPERTY
heiß	2b	PHYSICAL PROPERTY
warm	2b	PHYSICAL PROPERTY
kalt	2b	PHYSICAL PROPERTY
hell	2b	PHYSICAL PROPERTY
laut	2b	PHYSICAL PROPERTY
leise	2b	PHYSICAL PROPERTY
kühl	2b	PHYSICAL PROPERTY
langsam	2b	PHYSICAL PROPERTY
still	2b	PHYSICAL PROPERTY
düster	2b	PHYSICAL PROPERTY
lahm	2b	PHYSICAL PROPERTY
alt	2b	AGE
neu	2b	AGE
jung	2b	AGE
unterschiedlich	3a	CONFORMITY
normal	3a	CONFORMITY
ähnlich	3a	CONFORMITY
besonders	3a	CONFORMITY
verschieden	3a	CONFORMITY
fremdartig	3a	CONFORMITY

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
kompliziert	3a	n.d.
schwierig	3a	n.d.
krank	3a	PHYSICAL STATE
gefährlich	3a	PHYSICAL PROPERTY
gesund	3a	PHYSICAL STATE
labil	3a	PHYSICAL PROPERTY
empfindlich	3a	PHYSICAL PROPERTY
schlapp	3a	PHYSICAL STATE
besorgt	3a	MENTAL STATE
zornig	3a	MENTAL STATE
zufrieden	3a	MENTAL STATE
töricht	3a	MENTAL STATE
aufgeregt	3a	MENTAL STATE
empört	3a	MENTAL STATE
traurig	3a	MENTAL STATE
sauer	3a	MENTAL STATE
grantig	3a	MENTAL STATE
enttäuscht	3a	MENTAL STATE
unzufrieden	3a	MENTAL STATE
beleidigt	3a	MENTAL STATE
glücklich	3a	MENTAL STATE
böse	3a	MENTAL STATE
ängstlich	3a	MENTAL STATE
spannend	3a	VALUE
aufregend	3a	VALUE
eklig	3a	VALUE
langweilig	3a	VALUE
komisch	3a	VALUE
öde	3a	VALUE
doof	3a	VALUE
merkwürdig	3a	VALUE
interessant	3a	VALUE
peinlich	3a	BEHAVIOR
vorsichtig	3a	BEHAVIOR
zappelig	3a	BEHAVIOR
übermütig	3a	BEHAVIOR
scheu	3a	BEHAVIOR

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
beschäftigt	3a	BEHAVIOR
schüchtern	3a	BEHAVIOR
lustig	3a	BEHAVIOR
witzig	3a	BEHAVIOR
hilfsbereit	3a	BEHAVIOR
wild	3a	BEHAVIOR
frech	3a	BEHAVIOR
einfach	3b	n.d.
schön	3b	PHYSICAL PROPERTY
schick	3b	PHYSICAL PROPERTY
elegant	3b	PHYSICAL PROPERTY
hässlich	3b	PHYSICAL PROPERTY
hübsch	3b	PHYSICAL PROPERTY
clever	3b	MENTAL STATE
schlau	3b	MENTAL STATE
froh	3b	MENTAL STATE
dumm	3b	MENTAL STATE
gut	3b	VALUE
lecker	3b	VALUE
toll	3b	VALUE
schlecht	3b	VALUE
schlimm	3b	VALUE
blöd	3b	VALUE
nett	3b	BEHAVIOR
fleißig	3b	BEHAVIOR
gnadenlos	3b	BEHAVIOR
anständig	3b	BEHAVIOR
herzlich	3b	BEHAVIOR
freundlich	3b	BEHAVIOR
artig	3b	BEHAVIOR
lieb	3b	BEHAVIOR
zärtlich	3b	BEHAVIOR
erst-	n.d.	ORDINAL NUMBER
zweit-	n.d.	ORDINAL NUMBER
dritt-	n.d.	ORDINAL NUMBER
neunzehnt-	n.d.	ORDINAL NUMBER
zwanzigst-	n.d.	ORDINAL NUMBER

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
fünft-	n.d.	ORDINAL NUMBER
sechzehnt-	n.d.	ORDINAL NUMBER
sechst-	n.d.	ORDINAL NUMBER
viert-	n.d.	ORDINAL NUMBER
dreizehnt-	n.d.	ORDINAL NUMBER
achtzehnt-	n.d.	ORDINAL NUMBER
siebzehnt-	n.d.	ORDINAL NUMBER
hinter-	n.d.	POSITION
vorder-	n.d.	POSITION
ober-	n.d.	POSITION
unter-	n.d.	POSITION
recht-	n.d.	POSITION
link-	n.d.	POSITION
mittig-	n.d.	POSITION
mittler-	n.d.	POSITION
einzig	n.d.	QUANTITY
gesamt-	n.d.	QUANTITY
halb-	n.d.	QUANTITY
doppelt-	n.d.	QUANTITY
ganz	n.d.	QUANTITY
komplett	n.d.	QUANTITY
einsam	n.d.	QUANTITY
allein	n.d.	QUANTITY
vollständig	n.d.	QUANTITY
restlich	n.d.	QUANTITY
einzeln	n.d.	QUANTITY
nächst-	n.d.	TEMPORAL PROPERTY
letzt-	n.d.	TEMPORAL PROPERTY
ander-	n.d.	CONFORMITY
riesig	n.d.	DIMENSION
riesengroß	n.d.	DIMENSION
winzig	n.d.	DIMENSION
glasig	n.d.	PHYSICAL PROPERTY
futsch	n.d.	PHYSICAL PROPERTY
distelig	n.d.	PHYSICAL PROPERTY
piekselig	n.d.	PHYSICAL PROPERTY
zuckerkrank	n.d.	PHYSICAL PROPERTY

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
wunderschön	n.d.	PHYSICAL PROPERTY
uralt	n.d.	AGE
wunderbar	n.d.	VALUE
super	n.d.	VALUE
geil	n.d.	VALUE
zauberhaft	n.d.	VALUE
klasse	n.d.	VALUE
prima	n.d.	VALUE
dufte	n.d.	VALUE
fürchterlich	n.d.	VALUE
steil	n.d.	n.d.
arm	n.d.	n.d.
fertig	n.d.	n.d.
fein	n.d.	n.d.
eigen-	n.d.	n.d.
ruhig	n.d.	n.d.
glitzerig	n.d.	n.d.
frisch	n.d.	n.d.
mickrig	n.d.	n.d.
bereit	n.d.	n.d.
windig	n.d.	n.d.
übrig	n.d.	n.d.
tiptop	n.d.	n.d.
bequem	n.d.	n.d.
raffiniert	n.d.	n.d.
wichtig	n.d.	n.d.
spät	n.d.	n.d.
eilig	n.d.	n.d.
knapp	n.d.	n.d.
fett	n.d.	n.d.
modern	n.d.	n.d.
lütt	n.d.	n.d.
chaotisch	n.d.	n.d.
reif	n.d.	n.d.
königlich	n.d.	n.d.
schnittig	n.d.	n.d.
brüllerig	n.d.	n.d.

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
mühselig	n.d.	n.d.
saftig	n.d.	n.d.
berühmt	n.d.	n.d.
roh	n.d.	n.d.
goldig	n.d.	n.d.
neblig	n.d.	n.d.
frei	n.d.	n.d.
stolz	n.d.	n.d.
fahrbereit	n.d.	n.d.
kernig	n.d.	n.d.
passend	n.d.	n.d.
putzig	n.d.	n.d.
praktisch	n.d.	n.d.
besetzt	n.d.	n.d.
anstrengend	n.d.	n.d.
ordentlich	n.d.	n.d.
bekannt	n.d.	n.d.
irr	n.d.	n.d.
schleierhaft	n.d.	n.d.
heikel	n.d.	n.d.
staubig	n.d.	n.d.
kitschig	n.d.	n.d.
riesen	n.d.	n.d.
verflixt	n.d.	n.d.
listig	n.d.	n.d.
endlos	n.d.	n.d.
feierlich	n.d.	n.d.
möglich	n.d.	n.d.
steil	n.d.	n.d.
kräftig	n.d.	n.d.
grob	n.d.	n.d.
speziell	n.d.	n.d.
natürlich	n.d.	n.d.
ausziehbar	n.d.	n.d.
elektrisch	n.d.	n.d.
gefräßig	n.d.	n.d.
brummig	n.d.	n.d.

Table A.1: Leo: Complete list of adjectives produced between 2;00 and 2;11 years ordered according to semantic classes and grouped by notional class.

Adjective	Semantic class	Notional class
nadelig	n.d.	n.d.
gierig	n.d.	n.d.
zäh	n.d.	n.d.
doppelstöckig	n.d.	n.d.
babymäßig	n.d.	n.d.
schadhaft	n.d.	n.d.

Note. 1 = subsective/intersective, non-gradable, 2a = subsective/intersective, absolute gradable dimensional, 2b = subsective/intersective, relative gradable dimensional, 3a = subsective/intersective, absolute gradable evaluative, 3b = subsective/intersective, relative gradable evaluative, n.d. = not defined.

A.2 Conceptual level: Development of tokens per notional class

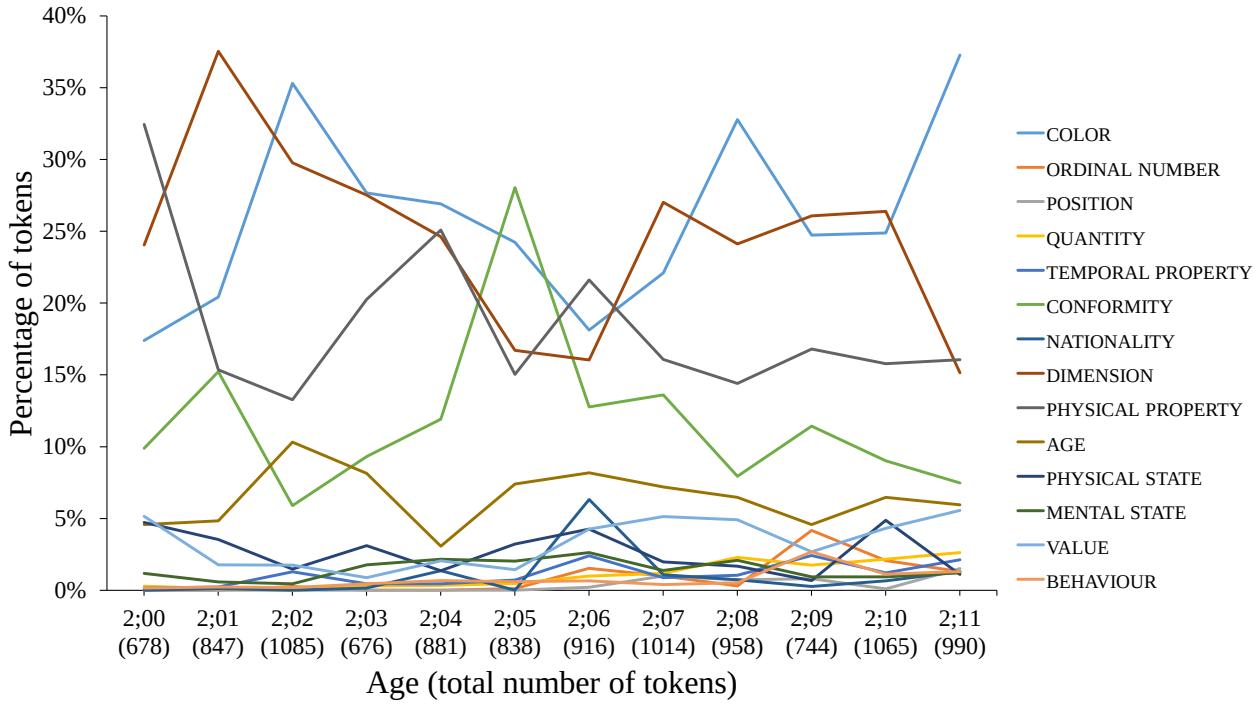


Figure A.1: Leo: Distribution of adjective tokens per notional adjective class across age.

Table A.2: Leo: Correlations between the total number of tokens (in %) and age per notional class.

	Age
Total number of tokens	
AGE	-.051
BEHAVIOR	.225
COLOR	.061
CONFORMITY	-.120
DIMENSION	-.028
MENTAL STATE	-.324*
NATIONALITY	-.297
PHYSICAL PROPERTY	-.111*
PHYSICAL STATE	-.343*
VALUE	.021

Note. * $p < .05$.

A.3 Conceptual level: Development of types per notional class

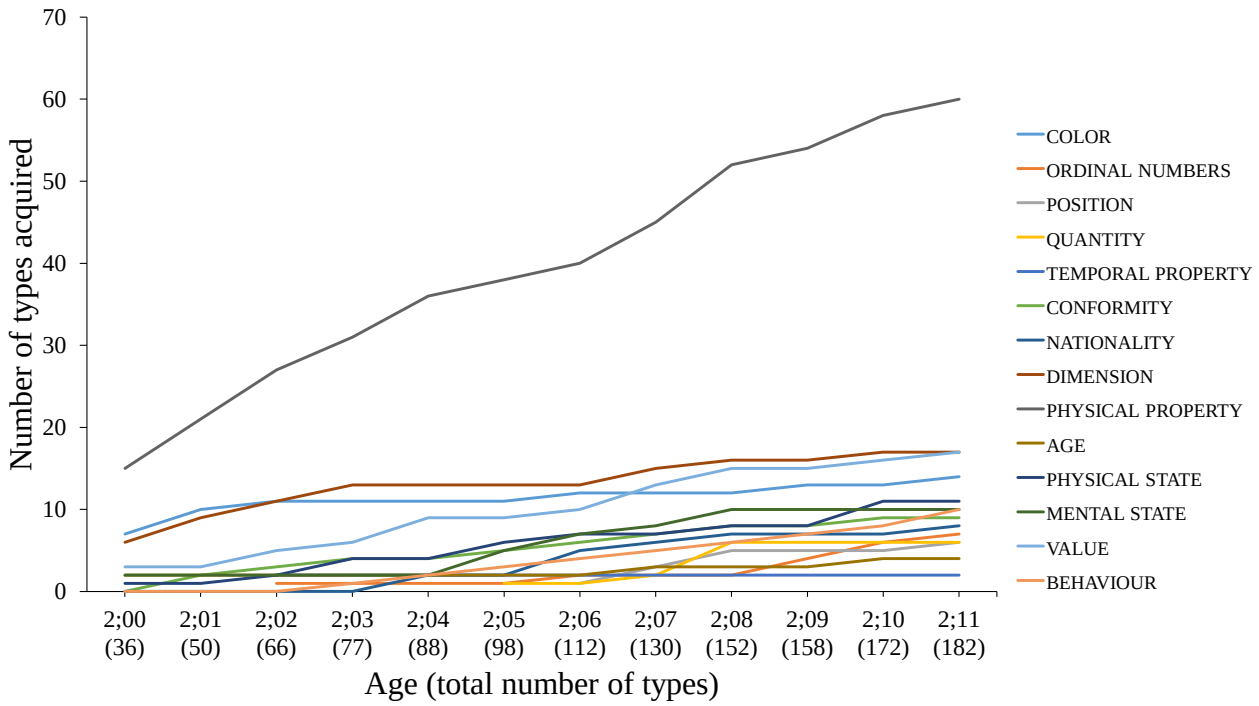


Figure A.2: Leo: Distribution of adjectives types per notional adjective class across age.

Table A.3: Leo: Correlations between the total number of types and age per notional class.

	Age
Total number of types	
AGE	.887***
BEHAVIOR	.993***
COLOR	.973***
CONFORMITY	.995***
DIMENSION	.979***
MENTAL STATE	.946***
NATIONALITY	.973***
ORDINAL NUMBERS	.973***
PHYSICAL PROPERTY	1.00***
PHYSICAL STATE	.991***
POSITION	.929***
QUANTITY	.944***
TEMPORAL PROPERTY	.648*
VALUE	.995***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

A.4 Conceptual level: Age of first and repeated use for notional classes

Table A.4: Leo: Age of first use in months for notional classes.

Adjective class	*N*	Age range	Median	Mean (*SD*)
AGE	4	24 - 31	24.00	25.75 (3.5)
BEHAVIOR	22	24 - 34	29.00	29.27 (3.0)
COLOR	16	24 - 35	24.00	26.00 (3.4)
CONFORMITY	12	24 - 33	28.00	28.08 (3.4)
DIMENSION	20	24 - 34	25.00	26.75 (3.4)
MENTAL STATE	19	24 - 35	30.00	29.84 (3.1)
NATIONALITY	11	25 - 35	30.00	30.45 (3.3)
ORDINAL NUMBERS	12	26 - 34	33.00	31.83 (2.4)
PHYSICAL PROPERTY	75	24 - 35	26.50	27.86 (3.6)
PHYSICAL STATE	14	24 - 35	29.00	28.43 (3.6)
POSITION	8	30 - 32	31.00	31.25 (0.7)
QUANTITY	10	24 - 34	31.00	29.90 (3.3)
TEMPORAL PROPERTY	2	24 - 25	24.50	24.50 (0.7)
VALUE	23	24 - 35	28.00	28.61 (3.7)

Note. N = Number of types, SD = Standard deviation.

Table A.5: Leo: Age of repeated use in months for notional classes.

Adjective class	*N*	Age range	Median	Mean (*SD*)
AGE	4	24 - 34	27.50	28.25 (5.1)
BEHAVIOR	10	27 - 35	31.50	31.40 (2.9)
COLOR	14	24 - 35	24.50	26.21 (3.7)
CONFORMITY	9	24 - 34	29.00	28.67 (3.4)
DIMENSION	17	24 - 34	25.00	26.65 (3.3)
MENTAL STATE	10	24 - 32	29.50	29.00 (2.9)
NATIONALITY	8	28 - 35	30.00	30.50 (2.3)
ORDINAL NUMBERS	7	26 - 35	33.00	32.14 (3.1)
PHYSICAL PROPERTY	60	24 - 35	27.00	28.05 (3.6)
PHYSICAL STATE	11	24 - 34	29.00	29.64 (3.5)
POSITION	6	30 - 35	31.50	31.83 (1.7)
QUANTITY	7	31 - 33	32.00	32.00 (0.6)
TEMPORAL PROPERTY	2	26 - 26	26.00	26.00 (0.0)
VALUE	17	24 - 35	28.00	28.88 (3.5)

Note. *N* = Number of types, *SD* = Standard deviation.

A.5 Semantic level: Age of first use for semantic classes

Table A.6: Leo: Age of first use in months for semantic classes.

Adjective class	*N*	Age range	Median	Mean (*SD*)
1 'BLUE'	38	24 - 35	28.00	28.37 (3.9)
2a 'CLEAN'	50	24 - 35	26.00	27.20 (3.2)
2b 'BIG'	39	24 - 35	25.00	26.77 (3.5)
3a 'HEALTHY'	50	24 - 25	30.00	29.38 (2.9)
3b 'BEAUTIFUL'	25	24 - 35	28.00	28.48 (3.5)

Note. *N* = Number of types, *SD* = Standard deviation.

Table A.7: Leo: Age of first attributive use in months for semantic classes.

Adjective class	*N*	Age range	Median	Mean (*SD*)
1 'BLUE'	26	24 - 35	26.50	27.85 (3.4)
2a 'CLEAN'	25	24 - 35	29.00	29.08 (3.5)
2b 'BIG'	28	24 - 35	26.50	27.89 (3.5)
3a 'HEALTHY'	19	25 - 35	31.00	30.95 (3.1)
3b 'BEAUTIFUL'	12	24 - 34	29.00	28.92 (3.5)

Note. *N* = Number of types, *SD* = Standard deviation.

Table A.8: Leo: Age of first predicative use in months for semantic classes.

Adjective class	*N*	Age range	Median	Mean (*SD*)
1 'BLUE'	20	25 - 35	28.00	28.70 (2.9)
2a 'CLEAN'	37	24 - 35	29.00	29.38 (2.8)
2b 'BIG'	34	24 - 35	27.00	28.79 (2.8)
3a 'HEALTHY'	36	26 - 35	29.00	29.61 (2.3)
3b 'BEAUTIFUL'	18	25 - 35	29.00	29.17 (2.9)

Note *N* = Number of types, *SD* = Standard deviation.

Appendix B

Study 2

B.1 Practice items

Gib mir bitte die Fußbälle. ('Please give me the soccer balls.')

Gib mir bitte die Puppen. ('Please give me the dolls.')

Gib mir bitte die Eimer. ('Please give me the buckets.')

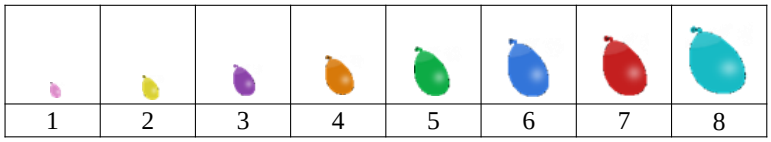
B.2 Items Part 1

B.2.1 Test items

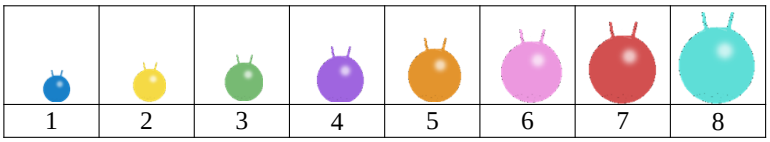
The objects are presented in linear order for illustration purposes only.

Items for relative gradable adjectives:

Session 1: *Gib mir bitte die großen Hüpfälle.*
Session 2: *Gib mir bitte die kleinen Hüpfbälle.*
(`Please give me the big/small space hoppers.')

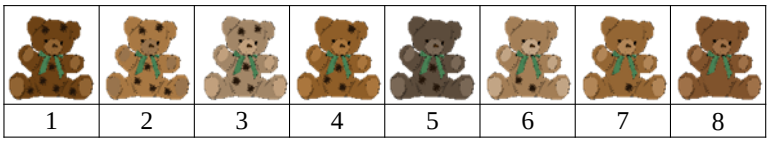


Session 1: *Gib mir bitte die großen Wasserbomben.*
Session 2: *Gib mir bitte die kleinen Wasserbomben.*
(`Please give me the big/small water balloons.')

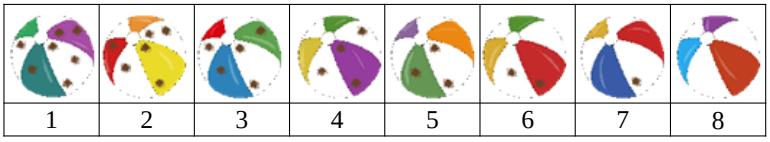


Items for absolute gradable adjectives:

Session 1: *Gib mir bitte die sauberen Teddys.*
Session 2: *Gib mir bitte die dreckigen Teddys.*
(`Please give me the clean/dirty teddies.')

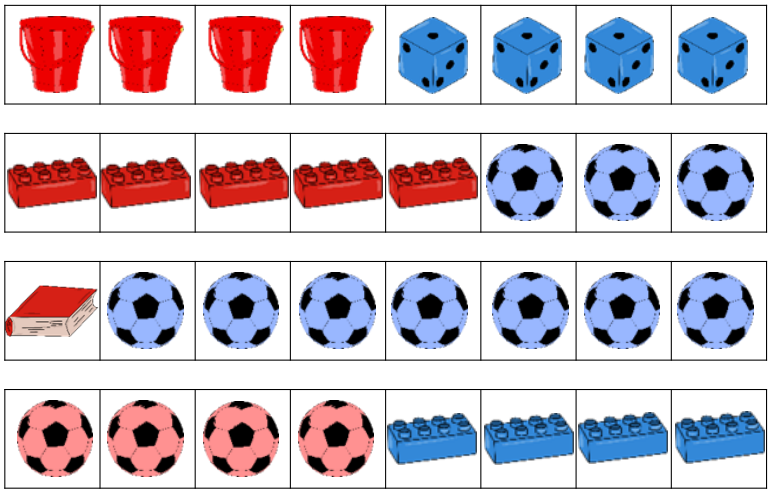


Session 1: *Gib mir bitte die sauberen Bälle.*
Session 2: *Gib mir bitte die dreckigen Bälle.*
(`Please give me the clean/dirty balls.')



B.2.2 Filler items

Visual contexts for the filler items:



B.3 Additional results Part 1

Table B.1: Statistical comparison (Wilcoxon Signed-Ranks Test) of the cut-off points for each test trial per age group.

Age		*big water balloons-big space hoppers*	*small water balloons-small space hoppers*	*clean teddies-clean balls*	*dirty teddies-dirty balls*
3	*Z*	-1.035	-.447	.000	.000
	p	.301	.655	1.000	1.000
4	*Z*	-.214	-.811	.000	.000
	p	.831	.417	1.000	1.000
5	*Z*	-.632	-2.530	.000	.000
	p	.527	.011	1.000	1.000
Adults	*Z*	-2.066	-1.633	.000	.000
	p	.039	.102	1.000	1.000

Table B.2: Mean and median cut-off point for each trial per child and adult group, and significance according to Mann-Whitney-U tests.

Trial	Children		Adults		Mann-Whitney-U
	Mean	Median	Mean	Median	*p*
big water balloons	5.2	5.0	5.9	6.0	.088
big space hoppers	5.3	5.0	6.2	6.0	.003
small water balloons	3.1	3.0	3.0	3.0	.648
small space hoppers	3.1	3.0	3.2	3.0	.818
clean teddies	7.8	8.0	8.0	8.0	.095
clean balls	7.4	8.0	8.0	8.0	.048
dirty teddies	7.0	7.0	7.0	7.0	.462
dirty balls	6.9	7.0	7.0	7.0	1.00

Table B.3: Mean and median cut-off point for each trial per child group, and significance according to Kruskal-Wallis tests.

Trial	3-year-olds		4-year-olds		5-year-olds		Kruskal-Wallis
	Mean	Median	Mean	Median	Mean	Median	*p*
big water balloons	4.2	4.5	5.6	6.0	5.5	5.0	.127
big space hoppers	4.9	5.0	5.5	6.0	5.3	5.0	.598
small water balloons	3.1	3.0	3.1	3.0	3.0	3.0	.980
small space hoppers	3.2	3.0	2.8	2.5	3.4	4.0	.341
clean teddies	7.8	8.0	7.5	8.0	7.9	8.0	.444
clean balls	7.2	8.0	7.1	8.0	7.9	8.0	.298
dirty teddies	7.0	7.0	7.1	7.0	7.0	7.0	.420
dirty balls	6.5	7.0	7.1	7.0	7.0	7.0	.415

Table B.4: Number of choices for each object per trial in the group of 3-year-olds.

Trial	Object							
	1	2	3	4	5	6	7	8
big space hoppers	0	1	2	3	7	10	11	11
big water balloons	2	2	3	5	8	10	10	10
small space hoppers	10	10	6	5	1	0	1	0
small water balloons	11	11	7	3	1	0	0	0
clean teddies	1	1	1	1	1	1	4	10
clean balls	1	1	1	1	1	1	3	11
dirty teddies	11	10	10	11	10	10	9	1
dirty balls	10	11	10	10	10	9	9	0

Note. Total number of possible choices per object = 11.

Table B.5: Number of choices for each object per trial in the group of 4-year-olds.

Trial	Object							
	1	2	3	4	5	6	7	8
big space hoppers	0	0	2	4	7	10	14	15
big water balloons	0	0	2	3	4	12	15	15
small space hoppers	15	13	8	5	3	1	0	1
small water balloons	15	15	8	5	2	1	0	0
clean teddies	1	2	1	2	1	2	2	15
clean balls	2	2	2	2	2	2	2	15
dirty teddies	15	15	15	15	15	15	15	1
dirty balls	15	15	15	15	15	15	15	1

Note. Total number of possible choices per object = 15.

Table B.6: Number of choices for each object per trial in the group of 5-year-olds.

Trial	Object							
	1	2	3	4	5	6	7	8
big space hoppers	0	0	2	3	12	14	16	17
big water balloons	0	0	1	2	11	13	15	17
small space hoppers	17	16	13	9	2	0	0	0
small water balloons	17	15	11	7	1	0	0	0
clean teddies	0	0	0	0	1	0	1	17
clean balls	0	0	0	0	0	0	1	17
dirty teddies	17	17	17	17	17	17	17	0
dirty balls	17	17	17	17	17	17	16	1

Note. Total number of possible choices per object = 17.

Table B.7: Number of choices for each object per trial in the group of adults.

Trial	Object							
	1	2	3	4	5	6	7	8
big space hoppers	0	0	0	0	4	20	23	26
big water balloons	0	0	0	1	7	22	26	26
small space hoppers	25	24	23	10	2	1	1	0
small water balloons	26	24	22	6	1	0	0	0
clean teddies	0	0	0	0	0	0	0	26
clean balls	0	0	0	0	0	0	0	26
dirty teddies	26	26	26	26	26	26	26	0
dirty balls	26	26	26	26	26	26	26	0

Note. Total number of possible choices per object = 26.

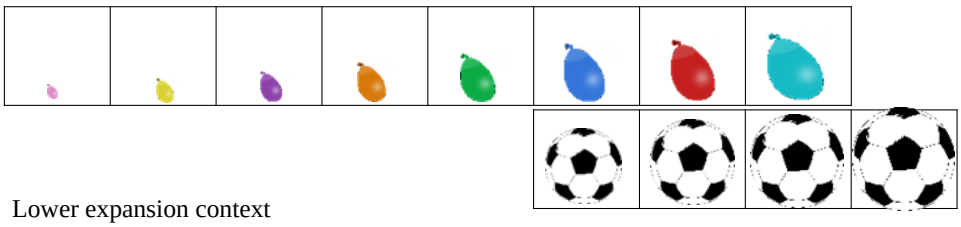
B.4 Items Part 2 and 3

B.4.1 Test items Part 2

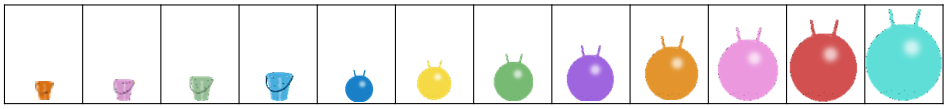
The objects are presented in linear order for illustration purposes only.

Items for relative gradable adjectives:

Upper expansion context
Session 1: *Gib mir bitte die großen Spielsachen.*
Session 2: *Gib mir bitte die kleinen Spielsachen.*
(`Please give me the big/small toys.')

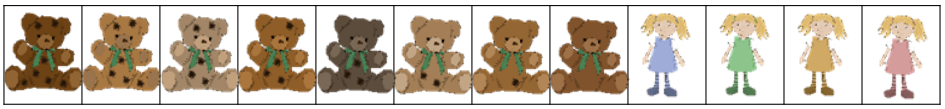


Lower expansion context
Session 1: *Gib mir bitte die großen Spielsachen.*
Session 2: *Gib mir bitte die kleinen Spielsachen.*
(`Please give me the big/small toys.')

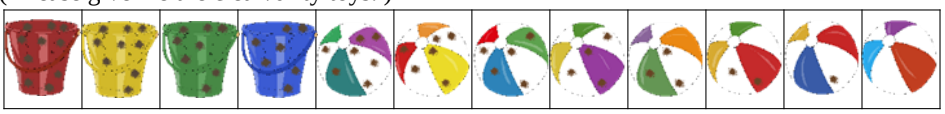


Items for absolute gradable adjectives:

Upper expansion context
Session 1: *Gib mir bitte die sauberen Spielsachen.*
Session 2: *Gib mir bitte die dreckigen Spielsachen.*
(`Please give me the clean/dirty toys.')



Lower expansion context
Session 1: *Gib mir bitte die sauberen Spielsachen.*
Session 2: *Gib mir bitte die dreckigen Spielsachen.*
(`Please give me the clean/dirty toys.')

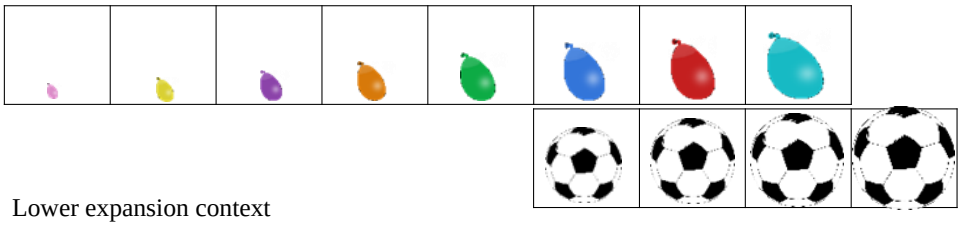


B.4.2 Test items Part 3

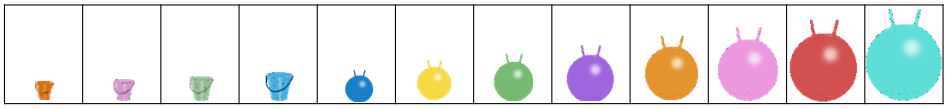
The objects are presented in linear order for illustration purposes only.

Items for relative gradable adjectives:

Upper expansion context
Session 1: *Gib mir bitte die großen Wasserbomben.*
Session 2: *Gib mir bitte die kleinen Wasserbomben.*
(`Please give me the big/small water balloons.')

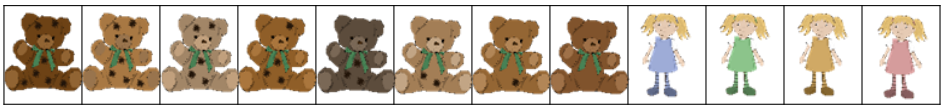


Lower expansion context
Session 1: *Gib mir bitte die großen Hüpfbälle.*
Session 2: *Gib mir bitte die kleinen Hüpfbälle.*
(`Please give me the big/small space hoppers.')

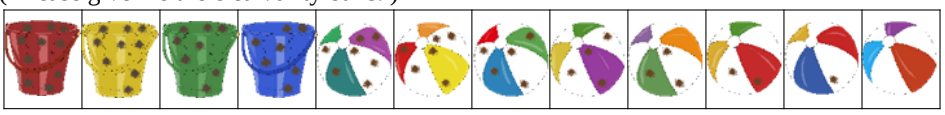


Items for absolute gradable adjectives:

Upper expansion context
Session 1: *Gib mir bitte die sauberen Teddys.*
Session 2: *Gib mir bitte die dreckigen Teddys.*
(`Please give me the clean/dirty teddies.')

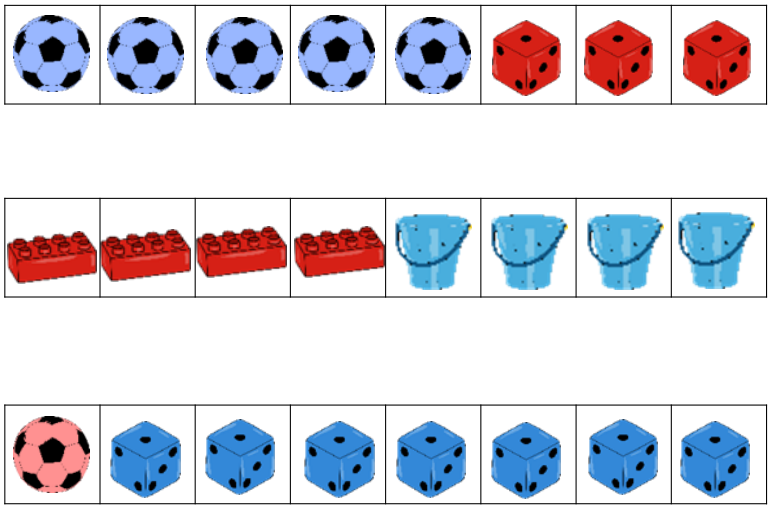


Lower expansion context
Session 1: *Gib mir bitte die sauberen Bälle.*
Session 2: *Gib mir bitte die dreckigen Bälle.*
(`Please give me the clean/dirty balls.')



B.4.3 Filler items

Visual contexts for the filler items:



B.5 Additional results Part 2 and 3

Because the participants had the chance to adjust the standard four times for RGs (twice for *big* = once in upper expansion context, once in lower expansion context; twice for *small* = once in upper expansion context, once in lower expansion context) and four times for AGs (twice for *clean* = once in upper expansion context, once in lower expansion context; twice for *dirty* = once in upper expansion context, once in lower expansion context), the total number of possible standard adjustments per adjective class was 44 in the 3-year-old group, 60 in the 4-year-old group, 68 in the 5-year-old group, and 104 in the adult group.

Table B.8: Number of standard adjustments per age group and adjective class.

Age	*N*	Relative gradable: *big/small* (n=4)	Absolute gradable: *clean/dirty* (n=4)
3	11	10	8
4	15	24	2
5	17	29	5
adults	26	47	1