Coherence and Cohesion in Early Immersion Students' L2 Narratives

Implications for Cognitive and Linguistic Development

Inauguraldissertation zur Erlangung des akademischen Grades „Doktor der Philosophie“ im Fach Englische Sprachwissenschaft an der Universität Paderborn

Erstgutachter: Prof. Dr. Manfred Pienemann

Zweitgutachter: Prof. em. Dr. Henning Wode

vorgelegt von

Christine Möller

Paderborn 2010

Acknowledgements

First of all, very special thanks to Henning Wode, without whom I would never have set foot on immersion territory, and to all teachers, students, and parents of Claus-Rixen-Schule!

There are too many other people to thank each one of them individually, but here are at least some of those who deserve a very special mention: Anna Zaunbauer, who read it all and, moreover, helped out whenever I had any statistics questions; her help was invaluable; Andreas Rohde, who was the second to read the manuscript in its entirety and gave important feedback; Claudia Claridge, who answered many major and minor questions over the years and always had an open ear for more questions. Jack Chambers, “the great encourager”, who read and commented on parts of the manuscript; Manfred Pienemann for helping out when help was most needed.

Many thanks also to the great many others who, in some way or other, contributed to the completion of this thesis or to making it possible in the first place: Sumru Akcan, Susanne Bendixen, Catherine Bennewitz, Petra Burmeister, Ping Deters, Johanna Gerwin, Christiane Gross, Swantje Hachmann, Kristin Kersten, Peter Kriwy, Normand Labrie, Christina Lupa, Matthias Meyer, Thorsten Piske, Ruth Pasternak, David Reuter, Bianca Sauer, Anja Steinlen, Stefanie Tacken, and Natalie Vellmer.

Last but not least thanks to my mother and especially to my husband Nour-Eddine for his almost endless patience, support, and encouragement.

Table of contents

List of figures

List of tables

1 Introduction 1

2 Narrative production: From cognition to coherence and cohesion 7

2.1 A simplified model of narrative discourse production 7

2.2 Top-down versus bottom-up organization: Narrative coherence and cohesion 15

2.3 The story grammar approach to narrative coherence 18

2.3.1 Story grammar 18

2.3.2 Story grammar: Narrative components 20

2.4 Cohesion: From references to lexical cohesion 24

2.4.1 References 25

2.4.2 Substitution and ellipsis 26

2.4.3 Connectives 27

2.4.4 Lexical cohesion 29

2.5 Narrative production: Summary and research questions 30

3 The development of storytelling 32

3.1 The development of narrative coherence in L1 acquisition 33

3.2 The development of narrative coherence in L2 and bilingual acquisition 36

3.3 The development of cohesion in L1 acquisition 38

3.4 The development of cohesion in L2 and bilingual acquisition 41

3.5 The relationship between coherence and cohesion 43

3.6 Summary 45

4 Research Design 48

4.1 Participants and data collection 48

4.1.1 The Kiel project 48

4.1.2 Participants 50

4.1.3 Data collection: Materials and procedure 55

4.2 Method of analysis 57

4.2.1 Narrative coherence: The structure of “Frog, where are you?” 57
4.2.2 Cohesion 64
4.2.3 Statistical methods 80
4.3 Addressing the dangers of the comparative fallacy 83

5 The development of coherence: Results 86
5.1 Total number of narrative components 86
5.1.1 Total number of components: Observed results 86
5.1.2 Total number of components: Statistical results 103
5.1.3 Summary: Total number of narrative components 104
5.2 Individual narrative components 106
5.2.1 Individual narrative components: Observed and statistical results 106
5.2.2 Summary: Individual narrative components 122
5.3 Index of global narrative structure 123
5.3.1 Index construction and reliability 124
5.3.2 Global narrative index results 125
5.3.3 Summary: Narrative index 137
5.4 Narrative coherence: Summary 138

6 The development of cohesion: Results 141
6.1 Overall cohesive density 141
6.1.1 Observed results overall cohesive density 141
6.1.2 Overall cohesive density by sex 144
6.1.3 Overall cohesive density by experience group 149
6.1.4 Statistical results overall cohesive density 155
6.1.5 Summary: Overall cohesive density 156
6.2 Cohesive density: Subcategories 157
6.2.1 References 158
6.2.2 Connectives 171
6.2.3 Substitution and ellipsis 185
6.2.4 Lexical cohesion 201
6.3 The subcategories of cohesion: Contribution to the overall density of participants’ stories 214
6.3.1 Overall contribution 214
6.3.2 Contribution by sex 216

6.3.3 Contribution by experience group 217
6.3.4 Qualitative changes 219
6.3.5 Summary: The subcategories' contribution to overall density 222
6.4 Cohesion results: Summary 222

7 The relationship between (the development of) coherence and cohesion 229

8 General discussion 234

9 Bibliography 252

10 Appendix 267
10.1 Individual cohesion scores ordered by frequency 267
10.2 Individual cohesion scores ordered by frequency and sex 269
10.3 Individual cohesion scores ordered by frequency and experience group . 271
10.4 Coherence and cohesion scores 273
10.4.1 Coherence scores: Overall measures 273
10.4.2 Coherence scores: Individual components 275
10.4.3 Cohesion scores 277

List of figures

Fig. 2.1 A simplified model of discourse production in a picturebook-elicited storytelling task. 8

Fig. 2.2 Narrative components of a single episode structure according to story grammar approaches 20

Fig. 4.1 The structure of the frog story in narrative components as identified by story grammars. 58

Fig. 5.1 Mean number of narrative components by grade as well as increase from first to fourth grade 87

Fig. 5.2 Distribution of the total number of component scores by grade 88

Fig. 5.3 Mean number of narrative components by sex and grade as well as increase from first to fourth grade 90

Fig. 5.4 Distribution of total number of narrative component scores by sex in first grade 92

Fig. 5.5 Distribution of the total number of components by sex in fourth grade 93

Fig. 5.6 Distribution of the total number of components by grade for male participants 94

Fig. 5.7 Distribution of the total number of components by grade for female participants 95

Fig. 5.8 Mean number of components by grade and experience as well as increase from first to fourth grade. 96

Fig. 5.9 Distribution of the total number of components by experience group in first grade. 97

Fig. 5.10 Distribution of the total number of components by experience group in fourth grade 100

Fig. 5.11 Distribution of the total number of components by grade for participants with exclusively monolingual preschool experience (mono group) 101

Fig. 5.12 Distribution of the total number of components by grade for children who attended a monolingual group in a preschool offering also bilingual groups (MB group) 101

Fig. 5.13 Distribution of the total number of components by grade for children with bilingual preschool experience (bili group) 102

Fig. 5.14 Frequencies of the 14 individual narrative components by grade 106

Fig. 5.15 Frequency of each narrative component in first grade in descending order 108

Fig. 5.16 Increase in frequency from first to fourth grade for each narrative component 109

Fig. 5.17 Frequency of each narrative component in fourth grade in descending order 110

Fig. 5.18 Frequency of each narrative component by sex in first grade 112

Fig. 5.19 Frequency of each narrative component by sex in fourth grade 115

Fig. 5.20 Frequency of each narrative component in first grade by experience group 117

Fig. 5.21 Frequency of each narrative component by experience group in fourth grade 120

Fig. 5.22 Mean narrative index score by grade as well as increase from first to fourth grade 126

Fig. 5.23 Distribution of narrative index scores by grade 127

Fig. 5.24 Mean narrative index score by grade and sex as well as increase from first to fourth grade 128

Fig. 5.25 Distribution of narrative index scores by sex in first grade 129

Fig. 5.26 Distribution of narrative index scores by sex in fourth grade 131

Fig. 5.27 Mean narrative index score by experience group and grade as well as increase from grade 1 to 4 132

Fig. 5.28 Distribution of narrative index scores by experience group in first grade. 133

Fig. 5.29 Distribution of narrative index scores by experience in fourth grade 136

Fig. 6.1 Mean cohesive density by grade as well as increase from first to fourth grade 142

Fig. 6.2 Distribution of participants' individual cohesive density scores by grade ... 143

Fig. 6.3 Mean cohesive density by grade as well as increase from first to fourth grade 144

Fig. 6.4 Distribution of participants' individual cohesive density scores by sex in first grade 146

Fig. 6.5 Distribution of participants' individual cohesive density scores by sex in fourth grade 147

Fig. 6.6 Distribution of participants' individual cohesive density scores by sex and grade 148

Fig. 6.7 Cohesive density by grade and experience group 149

Fig. 6.8 Distribution of individual cohesive density scores by experience group in first grade..... 151

Fig. 6.9 Distribution of participants' individual cohesive density scores by experience group in fourth grade..... 153

Fig. 6.10 Distribution of participants' individual cohesive density scores by experience group and grade..... 154

Fig. 6.11 Distribution of participants' individual reference density scores by grade 159

Fig. 6.12 Distribution of participants' individual reference density scores by sex in first grade..... 161

Fig. 6.13 Distribution of participants' individual reference density scores by sex and grade..... 162

Fig. 6.14 Distribution of participants' individual reference density scores by experience group in first grade..... 165

Fig. 6.15 Distribution of participants' individual reference density scores by experience group and grade 167

Fig. 6.16 Distribution of participants' individual connective density scores by grade 172

Fig. 6.17 Distribution of participants' individual connective density scores by sex in first grade..... 175

Fig. 6.18 Distribution of participants' individual connective density scores by sex and grade..... 176

Fig. 6.19 Distribution of participants' individual connective density scores by experience group in first grade..... 179

Fig. 6.20 Distribution of participants' individual connective density scores by experience group and grade 181

Fig. 6.21 Distribution of participants' individual ellipsis density scores by grade..... 187

Fig. 6.22 Distribution of participants' individual ellipsis density scores by sex in first grade..... 190

Fig. 6.23 Distribution of participants' individual ellipsis density scores by sex and grade..... 191

Fig. 6.24 Distribution of participants' individual ellipsis density scores by experience group in first grade 194

Fig. 6.25 Distribution of participants' individual ellipsis density scores by experience group and grade..... 195

Fig. 6.26 Distribution of participants' individual lexical density scores by grade 202

Fig. 6.27 Distribution of participants' individual lexical density scores by sex in first grade..... 204

Fig. 6.28 Distribution of participants' individual lexical density scores by sex and grade..... 205

Fig. 6.29 Distribution of participants' individual lexical density scores by experience group in first grade..... 208

Fig. 6.30 Distribution of participants' individual lexical density scores by experience group and grade..... 209

Fig. 6.31 The subcategories of cohesion relative to the overall cohesive density in first and fourth grade..... 215

Fig. 6.32 The subcategories of cohesion relative to the overall cohesive density as a function of sex and grade..... 216

Fig. 6.33 The subcategories of cohesion relative to the overall cohesive density as a function of experience group and grade..... 218

Fig. 6.34 The subcategories of lexical cohesion relative to overall lexical density by grade..... 220

Fig. 6.35 The subcategories of reference relative to overall reference density by grade..... 220

Fig. 6.36 The subcategories of connectives relative to overall connective density by grade..... 221

Fig. 6.37 Cohesive densities by grade as well as their increase from first to fourth grade..... 223

List of tables

Tab. 4.1 Descriptive statistics of the study's participants 52

Tab. 4.2 Participants of cohort 1 53

Tab. 4.3 Participants of cohort 5 (grade 4) 53

Tab. 4.4 Participants of cohort 8 (grade 1) 54

Tab. 4.5 Reference types included in the analysis 73

Tab. 4.6 Subtypes of connectives included in the analysis 76

Tab. 4.7 Synonymy and near-synonymy included in the analysis 77

Tab. 4.8 Hyponymy included in the analysis 77

Tab. 4.9 Oppositional relations included in the analysis 78

Tab. 4.10 Word fields included in the analysis 78

Tab. 5.1 Descriptive statistics total number of narrative components by grade 87

Tab. 5.2 Distribution of the total number of component scores by grade 88

Tab. 5.3 Descriptive statistics total number of components by sex and grade 91

Tab. 5.4 Distribution of the total number of component scores by sex and grade 92

Tab. 5.5 Descriptive statistics total number of narrative components by experience group and grade 97

Tab. 5.6 Distribution of the total number of components by experience group and grade 98

Tab. 5.7 Frequencies of the 14 individual narrative components by grade 107

Tab. 5.8 Order of difficulty for first and fourth graders 111

Tab. 5.9 Frequency of each narrative component by sex and grade 113

Tab. 5.10 Order of difficulty for first graders by sex. 114

Tab. 5.11 Order of difficulty for fourth graders by sex 116

Tab. 5.12 Frequency of each narrative component in first and fourth grade by experience group 117

Tab. 5.13 Order of difficulty for first graders by experience group 119

Tab. 5.14 Order of difficulty for fourth graders by experience groups. 121

Tab. 5.15 Index construction: Components and their assigned weight 125
Tab. 5.16 Descriptive statistics narrative index by grade 126
Tab. 5.17 Distribution of narrative index scores by grade 127
Tab. 5.18 Descriptive statistics narrative index by sex and grade..... 129
Tab. 5.19 Distribution of narrative index scores by sex and grade..... 130
Tab. 5.20 Descriptive statistics narrative index by experience group and grade 132
Tab. 5.21 Distribution of narrative index scores by experience group and grade. .. 134
Tab. 6.1 Descriptive statistics cohesive density by grade 142
Tab. 6.2 Descriptive statistics cohesive density by grade and sex..... 145
Tab. 6.3 Descriptive statistics cohesive density by grade and experience group ... 150
Tab. 6.4 Descriptive statistics reference density 158
Tab. 6.5 Descriptive statistics reference density by sex and grade..... 160
Tab. 6.6 Descriptive statistics reference density by grade and experience..... 164
Tab. 6.7 Descriptive statistics connective density by grade 171
Tab. 6.8 Descriptive statistics connective density by sex and grade..... 174
Tab. 6.9 Descriptive statistics connective density by experience group and grade. 178
Tab. 6.10 Descriptive statistics ellipsis density by grade 186
Tab. 6.11 Descriptive statistics ellipsis density by sex and grade 189
Tab. 6.12 Descriptive statistics ellipsis density by experience group and grade..... 193
Tab. 6.13 Descriptive statistics lexical density by grade 201
Tab. 6.14 Descriptive statistics lexical density by sex and grade..... 203
Tab. 6.15 Descriptive statistics lexical density by experience group and grade..... 207
Tab. 6.16 Descriptive statistics cohesive densities by grade 224
Tab. 6.17 Interindividual variation in cohesive density by grade, sex, and experience group..... 225
Tab. 7.1 Correlations in first grade..... 230
Tab. 7.2 Correlations in mono first graders' data. 230
Tab. 7.3 Correlations in bili first graders' data..... 231

Tab. 7.4 Correlations in fourth grade 231
Tab. 7.5 Correlations in mono fourth graders' data 232
Tab. 7.6 Correlations in bili fourth graders' data 232
Tab. 7.7 Correlations among increase variables from first to fourth grade 233
Tab. 10.1 First graders' cohesive density scores in ascending order 267
Tab. 10.2 Fourth graders' cohesive density scores in ascending order 268
Tab. 10.3 Male first graders' cohesive density scores in ascending order 269
Tab. 10.4 Female first graders' cohesive density scores in ascending order 269
Tab. 10.5 Male fourth graders' cohesive density scores in ascending order 270
Tab. 10.6 Female fourth graders' cohesive density scores in ascending order 270
Tab. 10.7 Bili first graders' cohesive density scores in ascending order 271
Tab. 10.8 Mono first graders' cohesive density scores in ascending order 271
Tab. 10.9 Bili fourth graders' cohesive density scores in ascending order 272
Tab. 10.10 Mono fourth graders' cohesive density scores in ascending order 272

1 Introduction

> It is somewhat striking that so much research effort and interest has been focused on trying to understand how children come to learn the sounds, words, and syntax necessary to produce sentences in their native language, given that very little real language use is confined to the sentence level. (Pan & Snow 1999: 229)

Even though the situation has changed somewhat and a substantial body of research on phenomena above the sentence level is available today, the majority of these studies has been conducted on monolingual speakers. Far fewer studies have targeted bilinguals and L2 learners and even fewer have addressed learners in immersion (IM) programs. By looking at the development of narrative discourse in an English immersion program at elementary school in Germany the present study seeks to contribute to filling this research gap.

What is narrative discourse?

'Discourse' can be defined as the "use of language beyond a single sentence" (Bamberg & Moissinac 2002: 395). This definition encompasses not only written and oral mode but also a broad range of discourse types from 'conversation' to more specific genres such as, for example, 'narratives'. 'Discourse' used in this sense refers to the same phenomenon described by other authors as 'text'.[1] Halliday and Hasan, for example, define text as "any passage, spoken or written, of whatever length, that does form a unified whole" (1976: 1). In the present study 'discourse' and 'text' will be used synonymously in the sense conveyed by both quotations. However,

[1] However, yet other authors distinguish between discourse as a more dynamic and text as a static entity (e.g. Cutting 2002: 2, Johnstone 2002: 2, Hoey 1996, Clark 1994).

both terms will be employed as referring to 'extended discourse' (Pan & Snow 1999), i.e. a sequence larger than just a couple of sentences.[2]

Narrative as a particular discourse genre includes subgenres such as (fairytale/ make-believe/ fictional) 'stories' and 'personal narratives', i.e. narratives about personal experience. Consequently, 'narrative discourse' can be defined as a spoken or written piece of extended discourse associated with the particular discourse genre 'narrative'.[3]

Why discourse?

As the initial quote from Pan and Snow already indicated, discourse, i.e. "a continuous stretch of … LANGUAGE larger than a SENTENCE" (Crystal 2003: 141), is an important part of human communication. Accordingly, discourse features are included in all influential models of language competence or performance (e.g. Bachman 1990) and in all major language assessment frameworks, e.g. Cambridge Certificate of Proficiency in English, TOEFL or the Common European Framework of Languages (Council of Europe 2001).

What makes discourse special, though? The rules of grammar operate on a very local level, predominantly on clauses and sentences, and not usually across sentence borders. The rules of discourse production, on the other hand, operate (also) on larger stretches of spoken or written text. Since both rule systems function at different levels, local grammaticality is thus a priori unrelated to discourse requirements (cf. e.g. Givón 1995).[4] A series of sentences as in example 1.1 would more likely be accepted as discourse (here: a story), for example, than example 1.2, even if the morphosyntax is target-like in 1.2 but not in 1.1:

1.1 *Boy go school. Friend not nice, take bike. Boy cry, no bike.*

1.2 *The boy goes to school every day. He owns a bike. The boy's brother also has a bike.*

That is, discourse production – be it in an L1 or L2 – may involve the formation of grammatical sentences, but more importantly it requires discourse-specific abilities, e.g. command over the linguistic means to connect stretches of speech or

[2] For minimal texts consisting of one sentence cf. Halliday and Hasan (1976: 7).
[3] In the following, the terms 'narrative', 'narrative discourse' and 'narrative text' will be used interchangeably.
[4] However, the production of syntactically and morphologically target-like clauses and sentences is of course desirable to facilitate understanding.

writing and the cognitive abilities to pre-plan for this linguistic as well as for a content-related, structural connectedness (Berman 2001a).[5]

Especially in a school context discourse competence is a valuable asset. A variety of studies has shown the importance of oral discourse competence with respect to the acquisition of literacy, for example, and especially the expository written mode crucial for academic success (e.g. Gumperz et al. 1986, Scollon & Scollon 1986, Michaels & Collins 1986). Gumperz et al. (1986), for example, found that learning to use a decontextualized perspective, i.e. using text as co-text, is essential to the acquisition of the written mode used in academic settings.

Why narrative discourse?

In studying the development of discourse abilities of monolingual normal-developing or brain-damaged adults and children (e.g. Berman & Slobin 1994, Reilly et al. 1998, Joanette & Brownell 1990), the study of narratives has proved especially useful, since narratives occur from a relatively early age on in conversation[6] (e.g. Nelson 1986, cf. also ch. 3) due to their strong socio-cultural importance (e.g. Bamberg & Moissinac 2002, Stein & Policastro 1984). Thus, narratives are often considered the most important discourse genre (e.g. Bamberg & Moissinac 2002, Reilly et al. 1998).[7] At the same time it has been shown that especially narrative discourse competence is important for academic success (Scollon & Scollon 1986, Michaels & Collins 1986) and that some aspects of narrative discourse competence can even be indicative for later mathematical achievement (O'Neill et al. 2004).

The present study: Goals and outline

The importance of (oral) discourse competence is, of course, not limited to monolingual education. To the contrary, immersion students and other L2 learners face an even greater challenge than monolinguals when asked to produce discourse in their L2, since even for young learners there may be a gap between cognitive and linguistic skills. At the same time immersion has been found to have a positive effect especially on participants' L2 conversational skills and willingness to communicate

[5] The exact requirements are of course determined by the particular discourse genre.
[6] As opposed to other discourse types, e.g. expository texts, which are introduced only later in formal schooling.
[7] Fantasy and adventure stories in turn, such as the one used in the present study, are a subgenre of narrative discourse (e.g. Shapiro & Hudson 1997, Hicks 1991).

(e.g. Wode 2009: 38, Baker & MacIntyre 2003, Johnson & Swain 1997, Harley et al. 1990; cf. also Smit 2008, Lazaruk 2007, Genesee 1987). But how about the production of make-believe stories, which are considered a very challenging type of narrative discourse, since they require a largely autonomous construction of text? How does this type of discourse develop in an immersion program?

The present study investigates make-believe stories produced by sixty-six first and fourth graders (mean age 6;7 and 9;7) in an early partial immersion program in the north of Germany, in which all subjects besides German language arts are taught in English. Learner variables collected were grade (first vs. fourth), sex (male vs. female), and L2 preschool experience (monolingual German vs. German-English bilingual). Participants' stories were obtained through a picture-elicited oral storytelling task administered at the end of both school years. These stories were then analysed in terms of two main discourse features: Cohesion and coherence, i.e. the linguistic connectedness of stretches of speech, for example via references or ellipses, and content connectedness through a global organization structure following an underlying narrative schema. Besides investigating the development of cohesion and coherence from first to fourth grade this study explored differences attributable to participants' sex and/or preschool experience.

Besides contributing to the investigation of (the development of) narrative discourse produced by L2, and especially immersion L2 learners, the present study also has a more concrete goal in relation to the IM program its data was collected in: Even though bilingual education is a very old phenomenon in Europe, over time monolingual L1 education had come to be seen as "natural" in most European countries and it was only in the last third of the 20th century that this view started to shift again (cf. Möller 2009). Especially in the last 10 years immense progress has been made, which shows in Germany, for example, in an increase in immersion and other bilingual education programs as well as the introduction of at least some foreign language teaching in elementary school. Nevertheless, prejudices and reservations by parents and policy makers continue. This shows, for example, in recurring discussions about the importance of German as a national language (e.g. Spiegel Online 2008a, 2008b, Welt Online 2008). Thus, the present study has a threefold aim:

1) To investigate how linguistic and content organization of (narrative) discourse develop over the four-year duration of an early partial IM program
2) to relate this development to participants' cognitive and linguistic development
3) to relate the overall results to the effectiveness of the program.

These three goals will be pursued as follows: In chapter 2 I will present the theoretical background of my study. First of all, I will outline the challenges involved in telling a story from a picture book with the help of a simplified model of (narrative) discourse production. This model leads up to a description of the two fundamental dimensions of texts investigated in my study, namely coherence and cohesion. Coherence will be defined as a text's organization structure reflecting the use of a narrative schema as a production plan. Cohesion, on the other hand, will be defined as the use of linguistic means to connect stretches of discourse. By continuity, coherence is defined as a cognitive and cohesion as a linguistic measure. Both discourse measures will be described in detail and several research questions will be posed in relation to them, which are addressed in the subsequent results chapters.

In chapter 3 I will give an overview of the findings of previous studies with respect to narrative development. Studies on monolingual L1 as well as on L2 and bilingual learners will be presented, albeit in separate sections. The overview of prior research will show that coherence and cohesion, as they were defined in chapter 2, can be studied fruitfully (a) within my participants' age range and (b) in L2 data. Chapter 3 concludes with several research hypotheses regarding the questions raised in the previous chapter.

The study's research design will be presented in chapter 4. Thus, a short overview will be given of the immersion project in which my data was collected. Then I will describe the participants and the data collection. After that I will describe the method of analysis both for coherence and cohesion in detail. In the last section of this chapter I will address the dangers of the comparative fallacy with respect to my analysis.

In chapter 5 I will present the results obtained for narrative coherence with respect to the number of components produced and the individual narrative components identified in the task material. Then the construction of a newly created index of global narrative structure will be described and the results presented, which

were obtained with the help of this index. Chapter 5 concludes with a summary of the coherence results.

The results obtained for narrative cohesion will be given in chapter 6. First of all, I will present the results obtained for overall cohesive density and then the ones for the subcategories. After that the subcategories' degree of contribution to the overall cohesion of participants' stories will be described. Chapter 6 ends with a summary of the cohesion results.

Chapter 7 gives, first of all, the results of a correlation analysis investigating the relationship between coherence and cohesion. Additionally, the results obtained for the relationship between both measures' development from first to fourth grade are explored.

In the last chapter, chapter 8, I will summarize and discuss the results of my study. Thus, I will address similarities and differences between coherence and cohesion results. Then I will focus on two recurring themes of my study, namely the influence of the learner variables collected (grade, sex, and L2 preschool experience) and on interindividual differences. In the light of my results I will then discuss the construct validity of coherence and cohesion as a cognitive, respectively linguistic measure and their usefulness for studying L2 data. After that I will address some general limitations of the present study. Chapter 8 ends with the discussion which conclusions can be drawn from my results for the effectiveness of the program at hand as well as IM teaching in general.

2 Narrative production: From cognition to coherence and cohesion

In the following I will present a simplified model of narrative discourse production focusing on the task participants faced in the present study (cf. ch. 4.1.3), namely telling a story from a picturebook (ch. 2.1).[1] Then I will take a "text as a product" stance and focus on (narrative) text as a result of bottom-up and top-down organization, which I argue to correspond to two fundamental dimensions recognized for (narrative) texts, namely cohesion and coherence (ch. 2.2). After that I will outline the approach to the analysis of narrative coherence chosen for the present study, story grammar, which recognizes an underlying story organization with different nodes that are related through membership in a hierarchical network (ch. 2.3). Narrative coherence is further explained to reflect participants' cognitive development. Finally, I will outline the theoretical basis for analyzing the narrative cohesion of participants' texts (ch. 2.4), i.e. their use of linguistic devices to create semantic links between individual textual elements. Cohesion is further explained to reflect participants' linguistic development. I conclude this chapter by giving a short summary and stating the three main research questions pursued in my study (ch. 2.5).

2.1 A simplified model of narrative discourse production

General agreement exists that texts are embedded in a communicative situation involving a producer and a (potential) receiver[2]. Fig. 2.1 shows a simplified model of

[1] As stated in ch.1, the term 'narrative' is used for a particular discourse genre, which includes, for example, (fairytale/ make-believe/ fictional) stories and personal narratives. In the following, 'narrative', 'story' and 'narrative text/ discourse' will be used interchangeably.

[2] Receivers do not necessarily have to be present, e.g. in an asynchronic communicative situation such as between writer and reader.

the processing steps involved in telling a story from a picturebook ('content material') to a listener. Due to the focus of the present study on production, the receiver side is not depicted. Text processing by the receiver is largely analogous, however, to the producer's processing of the elicitation material described in the present section (cf. Eysenck 2001, Cutler & Clifton 2000, Kintsch 1998).

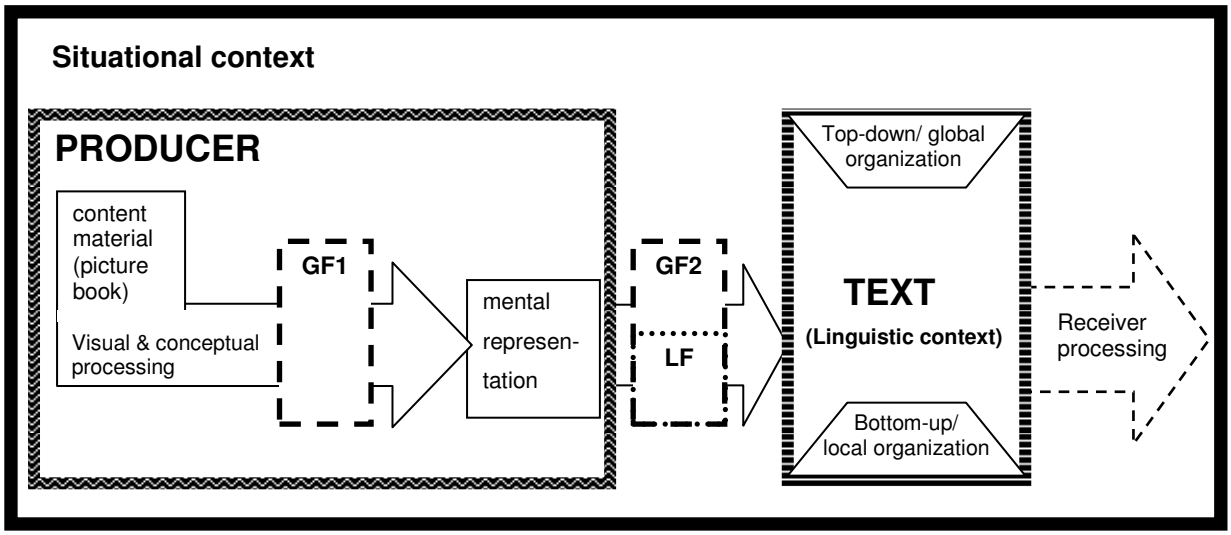

Fig. 2.1 A simplified model of discourse production in a picturebook-elicited storytelling task. Note: Processing repeatedly passes through a 'general filter' (GF1 & GF2), which corresponds to producers' knowledge base; in the encoding phase it additionally passes through a language-specific 'linguistic filter' (LF) (cf. text for further explanations).

The model focuses on the most important steps and processes, while leaving out many others, especially in regard to perceptual processing, linguistic en- or decoding and knowledge storage or access.[3] At the same time the model treats discourse production as a static, one-directional sequence of operations. In reality, a cyclic operation with constant feedback should be assumed due to limited working memory capacity (cf. Kintsch 1998, Rickheit et al. 1995), and several processes need to be considered largely parallel (cf. Eysenck 2001:3). Similarly, instead of a monologue an (at least partly) interactive process is closer to reality. This is also the case in the present study, where some interaction occurs, even if task and interviewer instructions deliberately aim for a monologue (cf. ch. 4.1.3).

As Fig. 2.1 shows, producers first of all need to process the content material cognitively in order to form a coherent mental representation of the content which can serve as a conceptual basis for the subsequent narration (e.g. Kintsch 1998 & 1995,

[3] For more information about perceptual and linguistic processing see, e.g., Herrmann & Fiebach 2004, Eysenck 2001, Cutler & Clifton 2000, Levelt 1999, Perfetti 1999. For additional information about memory storage and/ or access see, e.g., Schermer 2006: 116ff., Schneider & Büttner 2002, Eysenck 2001: 157ff. See also Foltz (2003) for an overview of existing models and challenges to modelling discourse processes.

Trabasso et al. 1995).[4] In this phase the producer corresponds to a 'receiver' or 'comprehender' of the elicitation material. Thus, producers need to process the content material perceptually, which corresponds to a 'bottom-up' or 'local' processing (e.g. Eysenck 2001: 382, Singer 1990: 109, Rumelhart 1980: 42).[5] At the same time all information from the content material is influenced by producers' 'knowledge base', i.e. any previous knowledge they bring to the task, which corresponds to a 'top-down' or 'global' processing of the content material (e.g. Eysenck 2001: 387, Kellogg 1995: 50f., Rumelhart 1980: 41f.).[6] The degree to which both processing modes (top-down and bottom-up) interact, i.e. the degree of parallel processing, depends on the experience with a task (Eysenck 2001: 3).

In each processing stage shown in the model, i.e. both in the producer-internal formation of a mental representation and in the encoding of this mental representation, producers' knowledge base is represented as a 'general filter' (GF1 & GF2). The model's GFs encompass 'general (world) knowledge' stored in long-term memory, 'situational knowledge' referring to the immediate communicative context as well as producers' immediate motivations and goals[7] (e.g. Eysenck 2001: 200, Perfetti 1999: 169, Kintsch 1995: 140, Kayser 1989: 346).

General knowledge comprises all of a person's previously acquired knowledge. It is based primarily on personal experience but also on mediating sources such as books or TV (e.g. Sodian 2002: 447ff., Nelson 1986a: 5).[8] General knowledge thus includes knowledge as diverse as theory of mind or conversational principles (e.g. Sodian 2002: 456ff., Levelt 1999: 84f., Grice 1975). Additionally, more abstract cognitive structures are formed through abstraction processes working on the representations of "real" experiences in memory: Schemata (e.g. Goswami 1998: 281, Singer 1990, Nelson 1986, Mandler & Goodman 1982).[9] Schemata in turn also become part of a person's general knowledge base.

---

[4] In telling a story from memory, on the other hand, producers would need to access the relevant memory structures.

[5] Also 'data-driven' or 'outside-in processing' (cf. Kellogg 1995: 50f., Anderson 1994, Singer 1990, Pearson 1982, Rumelhart 1980: 41f.).

6 Also 'hypothesis-driven', 'schema-based', 'conceptual', 'knowledge-driven' or 'inside-out processing' (cf. Kellogg 1995: 50f., Anderson 1994, Singer 1990: 108f., Pearson 1982, Rumelhart 1980: 41f.).

[7] Speakers' goals are not usually transparent. In the set-up of the present study, for example, the default speaker goal is assumed to be "follow the task instructions" (cf. ch. 4.1.3). The possibility remains, though, that speakers follow (additional) different goals such as "tell the story in time for the next break".

[8] For an alternative stance, i.e. innate theories, see Sodian (2002: 448).

[9] The terminology varies (cf. Kellogg 1995: 166), e.g. also 'cognitive model' (Ungerer & Schmid 2006) or 'global patterns' (Schubert 2008: 72, de Beaugrande & Dressler 1981: 90f.).

Situational knowledge (also 'contextual' knowledge), the second type of knowledge forming part of producers' knowledge base, includes 'interpersonal knowledge', i.e. knowledge about the interlocutor (the "target audience") and the social relation to her or him, as well as knowledge about the physical environment (Crystal 2003: 104). Situational knowledge cannot neatly be separated from the general knowledge base, however. General knowledge about social hierarchies in a specific culture is needed, for example, for the even more specific contextual knowledge about the relationship between interlocutors.

As explained above, schemata also form part of producers' knowledge base, which is represented by the model's GFs. In discourse production and reception schemata play important roles and they form the theoretical background for the analysis of narrative coherence conducted in the present study (cf. ch. 2.3). In the following I will therefore describe the role and characteristics of schemata in more detail before moving back to the description of the model.

A schema is, in its most general sense, "an organized knowledge structure" (Singer 1990: 6). Schemata range from semantic networks for simple objects, e.g. 'chair', to very complex entities such as stories (Schermer 2006: 161, Pearson 1982). Accordingly, different types of schemata have come to be distinguished, which can be embedded within each other. Best known are 'frames' and 'scripts' (cf. Kellogg 1995: 170ff., Rumelhart 1980, Schank & Abelson 1977, Minsky 1975). Frames refer to concept knowledge and stereotypical characteristics of concepts, e.g. STUDENT[10]. Scripts, on the other hand, refer to knowledge about people's stereotypical roles and activities in reference to certain situations, e.g. Schank's (1975) well-known RESTAURANT script. 'Schema' can thus be seen as a cover term with different subtypes, which corresponds to the usage adopted in the present study.

Schemata are organized in networks together with several subschemata. These networks are structured hierarchically in that more specific information is embedded under nodes[11] containing more general information (e.g. Ungerer & Schmid 2006: 50, Nelson 1986a: 8). A RESTAURANT script, for example, would incorporate

[10] Schemata will be written in capital letters.

[11] A variety of different terms is used to refer to schemata and their parts, the most general ones being 'variables' (e.g. Rumelhart 1980) or 'constituents' (e.g. Mandler & Goodman 1982, Thorndyke 1977). The term 'node' emphasizes the closeness to a connectionist model of the mind (Ungerer & Schmid 2006: 50). Other terms used are 'slots' (e.g. Pearson 1982, Minsky 1975), 'roles' (e.g. Schank & Abelson 1977, Schank 1975), and 'attributes' (cf. Schermer 2006: 148f., Smith 1991: 511).

ORDERING, PAYING, WAITER and several other subschema nodes (Schank 1975: 264ff.). All nodes are associated with default values, i.e. prototypical instances (e.g. Smith 1991: 510f., Pearson 1982: 28, Rumelhart 1980). A RUNNING RACE schema, for example, would contain a node PARTICIPANTS, and without information to the contrary, we would assume these participants to possess attributes such as 'human' and 'athletic'. However, schema nodes are also variable in nature, i.e. they have an attribute-value format (Smith 1991: 511): Once a particular schema is activated, the nodes' default values can be replaced by more specific ones on the basis of the incoming information (e.g. Pearson 1982, Rumelhart 1980). If contextual evidence stated, for example, that a turtle and a rabbit participated in the race, the default value for PARTICIPANTS would be substituted by turtle and rabbit.

Qualitatively, schemata are a function of the (amount of) experience they are based on (e.g. Johnson-Laird 1991: 483f., Singer 1990: 107f., Fivush & Slackman 1986). Consequently, schemata are person- as well as (sub-)culture-specific (e.g. Ungerer & Schmid 2006: 51ff., Scollon & Scollon 1986, Brewer 1985). Renkema (2004: 231) gives the HOUSE schema as an example: In Western culture, general agreement would be to include windows, different rooms, a kitchen, furniture etc., while this might not be the case in other cultures. At the same time, most people in Western culture would have different associations with the concept HOUSE, for example, a skyscraper or a farmhouse. Additionally, factors such as age, for example, may influence schemata even within the same culture. Children's STORYTELLING schema, for instance, may not necessarily correspond to that of adults, since early parent-child storytelling mainly consists in interactive picture description (Cook-Gumperz & Green 1986).

Schemata are involved in several ways in discourse comprehension, recall and production. First of all, schemata relevant to the content material, e.g. about prototypical events or story structures, are called up in producers' memory based on a bottom-up processing of the content material (e.g. Singer 1990: 108f., Pearson 1982: 28ff., Rumelhart 1980). In the set-up of the present study this includes not only the picturebook but also the task instructions (cf. ch. 4.1.3). The incoming information is then processed top-down, i.e. it is filtered as to correspondence with and relevance for existing schemata (e.g. Schermer 2006: 162, Singer 1990: 109, Pearson 1982: 30, Rumelhart 1980). That is, once a schema has been selected, the incoming information is processed in the light of this schema, seeking to attribute a specific

value to each of its (sub-)nodes. This process facilitates the mental organisation and storage of the information (e.g. Goswami 1998: 281, Kellogg 1995: 147ff.). In the present study, for example, participants were asked to look through a picturebook and then to "tell me [the interviewer] the story" (cf. ch. 4.1.3). Based on these task instructions, they must have expected, for example, that the booklet actually contains pictures. Similarly, they probably sought to identify one or more protagonists as well as a temporal or causal sequence of events – if they had accessed a relevant STORY schema. Participants would most certainly ignore, on the other hand, the paper quality or whether the interviewer was wearing a white or a blue shirt, since this information is irrelevant with respect to task and content material.

Additionally, schema-based processing enables inferences (e.g. Anderson 1994: 474, Mandler & Goodman 1982, Pearson 1982: 28).[12] This concerns 'bridging inferences' used to connect current and preceding parts of content material (e.g. Renkema 2004: 137f., Collins et al. 1980: 386), 'reconstructive inferences' (e.g. Anderson 1994: 474, Pearson 1982: 29), i.e. using default values in order to fill nodes that are not specified by the incoming information, as well as 'elaborative inferences', i.e. such that go beyond the information literally stated without being essential to the coherence of a message (e.g. Renkema 2004: 137f., Anderson 1994, Singer 1990: 175ff).[13] In the present study, for example, participants need to use schemata and other types of general knowledge to decode the contents of each individual picture and connect them to the content of previous and subsequent pictures. They need to infer, for example, that the frog escapes and that this escape triggers a reaction in the boy and the dog, namely to go looking for the frog (cf. ch. 4.2.1).

Finally, schemata also serve as production plans for the linguistic encoding on the part of the producer (e.g. Berman 2001b, Bamberg & Marchman 1991, Singer 1990: 109f., Kintsch & van Dijk 1978: 376). This involves different types of schemata at different levels of abstraction. A variety of different schemata would thus need to be activated in a study such as the present one, for example schemata relevant to the content material, such as GOING TO SLEEP or ANIMALS OF THE FOREST (cf. Wilkins in Berman & Slobin 1994: 21f.), or on a more abstract level a

[12] This is especially important since a large amount of information is left to be inferred and never realized explicitly in communication (Pearson 1982: 29).
[13] For further details on different types of inferences see, e.g., Singer 1990: 167ff., van de Velde 1989, Anderson & Pearson 1984.

STORYTELLING script specifying the details of the communicative situation. Very importantly, producers also need to activate a genre-specific STORY schema, which provides the abstract content components to be covered and their order, e.g. SETTING and ENDING (cf. ch. 2.3).

Evidence for the use of schemata as production plans comes primarily from recall experiments: In retelling a non-canonical story, for example, producers were found to restructure the original input and follow a classic STORY schema (Mandel & Johnson 1984). In another recall experiment, which involved a text with very general wording, participants were found to perform better if the global theme of the passage was known to them beforehand, since this allowed them to call up the relevant schemata for (a) input processing and storage and (b) for recall production (Bransford & Johnson 1972). These results are in agreement with the findings of Bartlett's (1932) groundbreaking study, which showed that participants in a recall experiment restructured the original content along schemata pertaining to their own cultural background.

To sum up, the model outlines that producers process the content material cognitively and "make sense" of it with the help of their knowledge base, i.e. the incoming information passes GF1, where any missing information is filled in. The result is a coherent mental representation on the part of the producer, which is a combination of content material and producers' additional knowledge sources (e.g. Kintsch 1995, Trabasso et al. 1995).[14] Up to this point, processing is identical to that of a receiver of (visual) information. In the next processing step, however, the producer's mental representation needs to be encoded linguistically. Schemata act as production plans for this linguistic encoding, i.e. the encoding is organized top-down.[15] In the model, the use of schemata as production plans is represented by processing passing again the general filter (GF2).

At the same time, producers also need to preplan their linguistic output with respect to receivers' needs, i.e. they need to try to predict the nature of the mental representation a receiver would form on the basis of the linguistic material (the text) received (e.g. Berman 2001b, Lenk 1998, Givón 1995, Kreyß 1995). Linguistic encoding can only begin if producers find that there is agreement between this

[14] In Kintschean terms, the 'situational model', i.e. a combination of 'text base' and additional knowledge sources (e.g. Kintsch 1995, van Dijk & Kintsch 1983).
[15] Cf. also Levelt's 'macroplanning' (1999: 91ff.).

presumed mental representation and their own. Otherwise, the planned linguistic output needs to be revised (Kreyß 1995).[16]

While all processing described so far (and thus its success) depends on general cognition, the encoding into speech (or writing) is language-specific, since different languages require different linguistic choices in the encoding process, even if one assumes the cognitive basis, i.e. the underlying mental representation, to be similar (e.g. Berman & Slobin 1994, Hickmann 1991).[17]

Linguistic encoding is conducted bottom-up[18] in combining sounds to words, words to sentences, sentences to discourse etc. (e.g. Sodian 2002: 483, Khodadady & Herriman 2000: 204ff., Rumelhart 1975: 235). In the model, the content of the mental representation is therefore shown as passing a language-specific 'linguistic filter' (LF), which contains knowledge about well-formed texts and their parts, i.e. knowledge about phonological structures, about the grammatical structure of words, clauses and sentences as well as knowledge about linguistic means to connect clauses, sentences, and longer stretches of speech. Consequently, successful encoding is dependent on producers' linguistic abilities in the respective language (be it an L1 or L2).

As the model indicates (cf. GF2 and LF), cognitive and linguistic processing are intimately connected in the encoding phase just as cognitive and perceptual processing were in the decoding phase (Eysenck 2001: 2ff., Rickheit et al. 1995, Bamberg 1987). Producers not only have to preplan, for example, but they also need to constantly monitor their own output against receiver needs and update their predictions about the mental representation formed by the receiver. Additionally, producers have to monitor their output against genre-specific schemata and other general, pragmatic and contextual requirements, which includes a constant awareness of the linguistic context already created (e.g. Berman 2001b, Lenk 1998, Hudson & Shapiro 1991).

As the model shows, the linguistic encoding of producers' mental representation finally results in a text (here: a 'story' or 'narrative'), which can then be processed by

[16] In reality, this is often done online, which is evident, for example, in self-repairs and comprehension checks such as *Do you know what I mean?*.
[17] As explained above, the intention is to present a very general processing model and thus no distinction is made between different types of linguistic processing modes such as vocabulary-driven versus grammar-driven (cf. the overview in Louwerse 2004).
[18] Cf. also Levelt's 'microplanning' (1999: 91ff.).

the receiver (here: the listener).[19] Texts as a "product" of the processing and encoding illustrated above also form the material for the analysis conducted in the present study. A "text as a product" view, i.e. disregarding any performance aspects, is common practice in studying discourse (cf. Berman & Slobin 1994a: 24). Nevertheless, it should be kept in mind that such a view of stories is a researcher-induced artificial reduction.

In the next section, the two text construction principles described above, bottom-up and top-down, will be set in relation to two fundamental dimensions of (narrative) discourse, namely coherence and cohesion, which are the research paradigms the present study is based on.

2.2 Top-down versus bottom-up organization: Narrative coherence and cohesion

Narrative texts share two fundamental dimensions with other types of text: Coherence and cohesion (e.g. Renkema 2004, Graesser et al. 2003, Hickmann 1996: 201, Gernsbacher & Givón 1995). Coherence and cohesion work together in making a text a unified, meaningful whole which is more than the sum of its sentences (Hoey 1996, Cook 1989, Halliday & Hasan 1976) and they correspond to the two different organization principles described in the last section, namely top-down organization of the content structure, and bottom-up organization of the linguistic structure (e.g. Karmiloff-Smith 1985: 62; cf. also ch. 2.1). There is of course more to stories and storytelling besides coherence and cohesion. Additional dimensions include, for example, a text's entertaining function, evaluative dimension or given-new structure (cf. Bamberg & Moissinac 2002, Brewer 1985, de Beaugrande & Dressler 1981). However, the present study focuses on coherence and cohesion as the two most fundamental dimensions of text.

The two most influential approaches in the study of narrative coherence have been highpoint analysis and story grammar (Peterson & McCabe 1983: 3).[20] Even though there is a certain similarity between the two approaches in that both analyze stories with the help of predetermined structural components, they also have important differences: Highpoint analysis (Labov 1999 [1972], Labov & Waletzky

[19] Again, this is a parallel process and the sequential description is owed to explanatory clarity.
[20] Other research paradigms stress, for example, form-function relationships (e.g. Berman & Slobin 1994, Bamberg & Marchman 1990) or interaction patterns (e.g. Quasthoff & Becker 2004). See also the overview in Bamberg & Moissinac 2002: 407ff.

1967) was developed from oral narratives about personal experience and it serves mainly sociolinguistic and anthropological purposes (e.g. Renkema 2004: 193, Toolan 1988: 46ff.). Story grammar, however, which is the approach to narrative coherence chosen for the present study, was developed on the basis of fairytale stories (e.g. Mandler & Johnson 1977, Thorndyke 1977, Bartlett 1932). Story grammar is a cognitive framework postulating a relationship between observed text structure and underlying cognitive structures. A more detailed overview of the theoretical assumptions underlying story grammar will be given in the following section (ch. 2.3).

Linguistically, texts are constructed bottom-up. That is, sounds are combined to words and words to sentences according to language-specific phonological, morphological and syntactic rules. Above the sentence level, however, i.e. in combining sentences to discourse, a discourse-specific linguistic rule system is at work: Cohesion. The notion of cohesion has been accepted as well defined and useful for the study of text since Halliday and Hasan's seminal work *Cohesion in English*, which was published in 1976 (Bublitz et al. 1999).

Cohesion is a textual, surface quality accomplished through the use of cohesive devices (e.g. Halliday & Hasan 1997 [1985] & 1976, Hickmann 1996: 201ff., Hoey 1996). Cohesive devices establish relationships of meaning, i.e. 'ties', between elements in the text by linking a presupposing with a presupposed element (Halliday & Hasan 1976: 4). These cohesive ties create 'texture', i.e. they linguistically establish the textual quality of a stretch of sentences (e.g. Shapiro & Hudson 1997, Halliday & Hasan 1976: 10).[21] Thus, cohesion can also be defined as "the set of possibilities that exist in the language for making text hang together" (Halliday & Hasan 1976: 18).

Cohesion is a phenomenon that applies to texts in general, so that no qualitative difference can be made between cohesive devices in narratives as opposed to other genres, even though there may be a genre-related influence on the specific choice and number of cohesive devices (Norment 1995). In strongly co-text dependent genres such as in the present study, for example, the use of cohesive devices is expected to be relatively high (Hickmann 2003: 43).

In their seminal work *Cohesion in English* Halliday and Hasan (1976) identify five main types of cohesive devices: References, connectives, substitution, ellipsis,

[21] As stated in ch. 1, Halliday and Hasan use the term 'text' for "any passage, spoken or written, of whatever length, that does form a unified whole" (1976: 1).

and lexical cohesion; these categories will be described in more detail in ch. 2.4. Following Halliday and Hasan's lead, several other linguistic variables were also found to contribute to cohesion, e.g. tense-aspect marking, punctuation, and intonation (cf. Hickmann 1996, Bamberg 1987, Gumperz et al. 1986, de Beaugrande & Dressler 1981). The contribution which these latter variables make to cohesion will not be discussed further, however, since the present study is based on the traditional categories established by Halliday and Hasan (cf. ch. 2.4 and 4.2.2.2).

No agreement exists about the relationship between coherence and cohesion, even if most researchers regard coherence as a superordinate quality and cohesion as contributing to it (e.g. Halliday & Hasan 1997 [1985] & 1976, Hellmann 1995, Cook 1989, Tannen 1986). Thus, it is argued that by indicating the relationship between elements in the text, cohesive ties contribute to textual coherence. Cognitively oriented theories go even further; they claim that cohesive devices serve as "processing instructions", i.e. as linguistic cues signaling the receiver to build a coherent mental representation and how to do this (e.g. Louwerse 2004, Knott & Sanders 1998, Segal & Duchan 1997, Givón 1995, Sanford & Moxey 1995). However, conflictive evidence exists about the amount and types of processing cues receivers need (Kamalski et al. 2008) and thus this line of research can only be pursued within a receiver-centered perspective.

At the same time there is agreement that cohesion markers are neither necessary (example 2.1) nor sufficient (example 2.2) for creating coherence, even though the occurrence of some cohesive ties is the norm in naturally occurring texts (e.g. Martin 2001: 44, Hoey 1996: 12, Hellmann 1995, Tannen 1986). Thus, example 2.1 would be considered a coherent, albeit short sequence of utterances, even if it contains virtually no cohesive devices. Example 2.2, on the other hand, contains a lexical tie (*bone – bones*) as well as a referential tie (*they – bones*) but this sequence of utterances would not easily be interpreted as coherent.

2.1 *Nora fastened the seat belt. At flight altitude the attendant offered beverages.*[22] *Right before landing an elderly lady started to panic.*

2.2 *The dog was eating a bone. Strong bones are important. They are white*.

[22] The example shows that coherence can also influence the use of cohesive devices: The use of the definite article in *the attendant* can be explained by activation of the FLYING ON A PLANE schema (cf. Ungerer & Schmidt 2006: 212).

A developmental analysis of narrative productions, as in the present study, therefore needs to be restricted to a description of the use of cohesive devices by text producers – without exploring these devices' potential for creating coherence or as processing cues.

2.3 The story grammar approach to narrative coherence[23]

2.3.1 Story grammar

Most researchers "agree in assuming the existence of a pre-linguistic representation of a general organization common to all narrative" (Gombert 1992: 146), i.e. a narrative schema. In fact, general schema theory is founded mainly on Bartlett's discovery that stories are encoded and memorized with the help of a culturally-determined narrative schema (1932; cf. also Schermer 2006: 160ff., Solso 2005: 305ff.).[24]

Just as other types of schemata (cf. ch. 2.1), narrative schemata are based on experience, i.e. they are "constructed through an inductive, inferential, and abstractive process, resulting from repeated exposure" (Stein & Policastro 1984: 124). Thus, story schemas are grounded in world knowledge, e.g. about cause and effect or about social interactions, as well as knowledge about different narrative genres based on experience with reading, being told and telling stories or even being taught prototypical story structures at school (cf. ch. 4.1.2) (e.g. Hudson & Shapiro 1991: 89, Stein 1982, Mandler & Johnson 1977). Consequently, story schemas are not only culture- but also genre-specific, i.e. they also vary among text genres (e.g. Shapiro & Hudson 1991: 960, Scollon & Scollon 1986, Brewer 1985).

Actual stories can be described as "surface realizations" of the underlying narrative schema, i.e. the schema's abstract nodes are filled with the specific content of a particular story (Nelson & Gruendel 1986: 40, Thorndyke 1977: 83). Story grammar, in turn, is a formal rule system (Mandler & Goodman 1982: 507), which was developed since the 1960s – analogously to generative grammar (Berman 1995:

[23] Unless stated differently, the term 'coherence' will in the following be used in the sense of 'narrative coherence' described above.

[24] Bartlett's work was at least the most influential one. The German psychologist Otto Selz would need to be credited with the first schema theory in the 20th century, published in 1922, but his work went largely unnoticed in Europe at the time due to his Jewish origins and it was noticed in the anglophone world only after 1945 (Lück & Miller 2005, Rieger 1987, Kintsch 1982: 319).

288) – in order to describe regularities in the surface structure of stories, which reveal the underlying story schema (e.g. Mandler & Johnson 1977, Thorndyke 1977, cf. also ch. 2.3.2). Story grammar thus operates under the following assumptions (Stein & Albro 1997, Stein & Glenn 1979, Mandler & Johnson 1977, Thorndyke 1977):

- Stories have some kind of internal structure, which is relatively stable within a culture and narrative genre
- This internal structure can be described as an abstract, hierarchical network of higher and lower-order components (nodes) realized linguistically in the surface structure of a text
- This network of nodes, in turn, reflects the mental organization of story components in processors, i.e. their narrative schema.[25]

Story grammar approaches have repeatedly shown the psychological validity of story schema, especially through recall and comprehension studies (e.g. Stein & Policastro 1984, Mandler & Goodman 1982; cf. also ch. 2.1). Story grammar research on narrative productions, on the other hand, is a comparatively recent approach (Bamberg & Moissinac 2002: 408) and it should be noted that schemata used for understanding and production are not necessarily the same. Schemata could be used successfully in comprehension, for example, while not being under operative control for production (e.g. Nelson 1986a: 12, Mandler & Johnson 1977).

Story grammar analysis is used in the present study to analyze the narrative coherence of the stories elicited, i.e. their content structure. Narrative coherence serves as an indicator for cognitive development, since the structure of participants' texts reflects their cognitive preplanning abilities in the sense of having command over a narrative schema. The successful encoding of schema nodes is directly related to participants' linguistic abilities, however, and differences in linguistic proficiency, e.g. between adults and children or between L1 and L2 speakers, can therefore influence any results obtained for narrative coherence (cf. ch. 2.1).[26] In the present study the linguistic threshold for considering component nodes realized was set relatively low in order to do justice to such a potential gap between cognitive and linguistic abilities; this will become clear in ch. 4.2.1.

[25] Just as memory structures in general, story schemata are hypothetical constructs, i.e. they are not accessible to direct observation (e.g. Schermer 2006: 12f.). Instead, conclusions are drawn from input/ output to processes and structures in the mind (e.g. Eikmeyer et al. 1995). See also the similar discussion on accessing competence in an L2 through performance measures (Ellis 1994: 12f.).

[26] Again, this relates to the fundamental methodological question whether underlying competences can be tapped through performance measures (Ellis 1994: 12f.; cf. also Chafe 1980).

2.3.2 Story grammar: Narrative components

Prototypical stories, as described by story grammars, have a problem-resolution structure, and they are goal-based, i.e. a problem arises, the protagonist[27] forms a goal to solve the problem, carries out one or several attempts to do so and in the end either succeeds or fails. This prototypical structure is operationalized with the help of narrative components: According to story grammar approaches to narrative coherence, a story has a deep structure consisting of a number of narrative components, which are either 'terminal nodes', i.e. states or events[28] directly encoded in the surface structure of the story, or higher order nodes, which have terminal nodes attached to them (e.g. Stein & Glenn 1979, Mandler & Johnson 1977, Thorndyke 1977; cf. also the previous section).

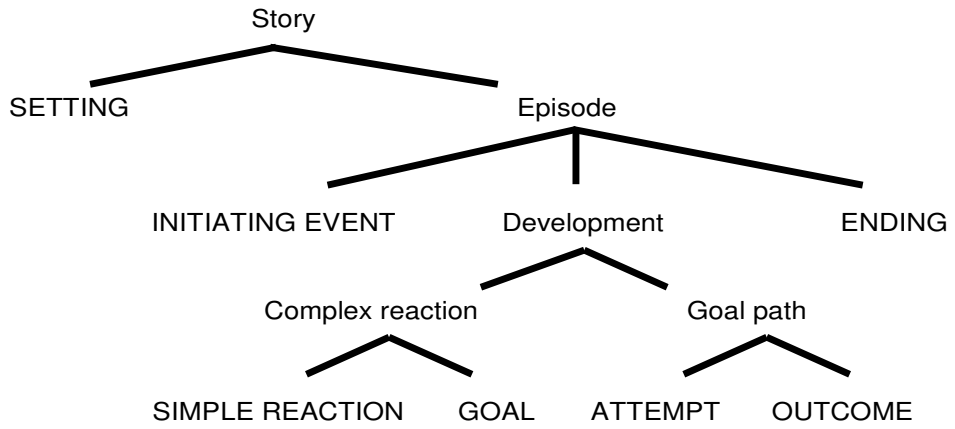

Fig. 2.2 Narrative components of a single episode structure according to story grammar approaches. Note: Terminal nodes are written in capital letters.

According to story grammars, a story minimally consists of a SETTING[29] component followed by an 'episode' (e.g. Westby 1984: 111, Stein & Glenn 1979: 59, Mandler & Johnson 1977: 117). A single episode includes an INITIATING EVENT, a DEVELOPMENT section, which incorporates the main body of the story, and an ENDING (see Fig. 2.2). The development section in turn consists of a complex reaction – caused by the initiating event – which can be subdivided into a SIMPLE REACTION and a GOAL-PLAN or simply GOAL. These are followed by a pursuit of the GOAL, i.e. the 'goal-path', which is composed of an ATTEMPT and its

[27] For better readability, I will be using the singular form 'protagonist' to stand for one or more protagonists, whether acting in unison or pursuing different goals.
[28] States and events can be either external or internal (cf. Peterson & McCabe 1983: 69f., Mandler & Johnson 1977: 115).
[29] In the following, terminal nodes will be written in capital letters.

OUTCOME. In the following, I will describe the terminal nodes of the story schema, as identified by most story grammars for a single episode, in more detail.

The SETTING[30] is one of the integral constituents of story grammars (e.g. Stein 1982, Mandler & Johnson 1977, Rumelhart 1975). It describes the state-of-affairs prior to the INITIATING EVENT, i.e. it provides information about initial time and location, participating characters and props, as well as additional background details (e.g. Shapiro & Hudson 1991, Mandler & Johnson 1977: 118). The minimal requirement for a setting is the mention of one animate character (Stein & Policastro 1984: 119, Mandler & Johnson 1977: 118), which is why the character introduction is also labeled 'major setting' as compared to other 'minor setting' information (Stein & Glenn 1979: 62).

The general function of the SETTING is to provide a contextual embedding for the listener[31] in which to integrate the ensuing episode (cf. Berman 2001b: 1, Peterson & McCabe 1991: 41). This listener-orientation requires the narrator to be aware of the contextual information a listener needs in order to build an initial representation in which he or she can successfully integrate the following episode(s) (cf. Berman 2001b: 2, Peterson & McCabe 1983: 54). In other words, the development of SETTING has to be seen in relationship with theory-of-mind development, i.e. the awareness of shared as opposed to unshared knowledge. As such, it forms part of children's cognitive development. Further evidence for this comes from recall studies: Character introduction, for example, which is the minimal requirement for a SETTING, seems to "act as a marker for the initiation of the [story] recall schema" (Stein & Glenn 1979: 62) just like beginning formulae such as *once upon a time* in English or *es war einmal* in German.

The INITIATING EVENT[32] (e.g. Berman & Slobin 1994, Stein & Policastro 1984: 118, Peterson & McCabe 1983: 67, Mandler & Johnson 1977: 119) marks the beginning of a story's problem-resolution structure. It consists of one or more events that make some change in the environment of the protagonist and typically triggers a complex reaction in the protagonist. Thus, the main function of the INITIATING

[30] In reference to highpoint analysis (Labov 1999 [1972], Labov & Waletzky 1967) sometimes also ORIENTATION.
[31] This study looks at oral narratives, thus the addressee is always referred to as a listener. The explanations given apply just the same to a reader.
[32] Also BEGINNING (e.g. Mandler & Goodman 1982) or ONSET of the plot (e.g. Berman & Slobin 1994).

EVENT is to evoke an emotional response in the protagonist (SIMPLE REACTION), which leads to the formation of the GOAL-PLAN.[33]

In a simple episode, INITIATING EVENTs are external or internal events, i.e. natural occurrences, action(s) performed by one of the story characters or thoughts about and perceptions of external events, which lead to changes in the habitual states described in the SETTING (Stein & Glenn 1979: 63, Mandler & Johnson 1977: 115). In more complex episodes, the INITIATING EVENT itself can consist of one or more subordinate episodes (e.g. Mandler & Johnson 1977: 115ff.).

The SIMPLE REACTION, which directly follows the initiating event, is often implicit, i.e. not overtly expressed in the story (Stein & Glenn 1979: 64, Mandler & Johnson 1977: 121). It refers to internal states of the protagonist in reaction to the INITIATING EVENT, i.e. any descriptions of emotional responses (e.g. sadness) or statements about a character's thoughts ('cognitions') (Shapiro & Hudson 1991, Stein & Glenn 1979: 64, Mandler & Johnson 1977: 119). The primary function of the SIMPLE REACTION is to motivate the GOAL-PLAN, which in turn results in the goal-path, i.e. action(s) by the protagonist (Stein & Policastro 1984: 144).

The GOAL-PLAN, or simply GOAL, consists of an internal plan of the protagonist to overcome the problem raised by the initiating event, i.e. statements referring to desires, intentions or strategies of the protagonist, which motivate the following plan application sequence, namely the goal-path (e.g. Trabasso & Rodkin 1994: 106, Shapiro & Hudson 1991, Stein & Glenn 1979). Just like the SIMPLE REACTION, the GOAL is often left to be inferred from the story context (e.g. Stein & Glenn 1979: 65, Mandler & Johnson 1977: 121).

The goal-path induced by the complex reaction consists of an ATTEMPT and an OUTCOME[34], i.e. one or several actions carried out in order to achieve the GOAL and the subsequent result(s) of these actions (e.g. Stein & Policastro 1984: 118, Mandler & Johnson 1977: 123). If the ATTEMPT is followed by a successful OUTCOME, i.e. the goal has been achieved and the problem resolved, the ENDING of the episode is reached (Stein & Glenn 1979: 58ff., Mandler & Johnson 1977: 123). The goal-path can be recursive, though: If the first attempt to reach the GOAL is marked by failure, the goal is reinstated[35] and motivates another attempt. Thus,

[33] Less typically, a SIMPLE REACTION is triggered, which is followed directly by an action (cf. Mandler & Johnson 1979: 119).
[34] Also CONSEQUENCE (e.g. Stein & Policastro 1984, Peterson & McCabe 1983).
[35] Reinstantiation can involve the creation of subgoals adapted to the specific circumstances described at that point in the narrative, i.e. local goal plans (Trabasso & Rodkin 1994: 93).

several attempts and their outcomes can follow each other until the overall goal is finally reached. As long as they are reinstantiations of the same main goal, these attempts are all part of the same ATTEMPT section (Trabasso & Rodkin 1994: 99ff., Mandler & Johnson 1977: 123, Thorndyke 1977: 80).

The ENDING[36] concludes an episode (e.g. Mandler & Goodman 1982: 509, Thorndyke 1977). Even though considerable redundancy between OUTCOME and ENDING may occur in stories with a single goal-path (cf. Mandler & Johnson 1977: 123), the two components can normally be clearly distinguished: While an OUTCOME is the immediate result of an ATTEMPT, the ENDING is connected to the overall GOAL and therefore to the episode as a whole (Mandel & Johnson 1984: 647, Mandler & Johnson 1977: 123). The ENDING wraps up the story and can include statements referring to one or several of the following: Successful attainment of the goal, emotional and cognitive responses of one or several story characters to the final state of affairs (i.e. goal attainment), actions resulting from these psychological states or long-term consequences occurring as a result of goal attainment (e.g. Stein & Policastro 1984: 118, Pradl 1979: 21, Mandler & Johnson 1977: 123). Long-term consequences can take the form of morals or genre-related linguistic formulae such as *they lived happily ever after*.

The task material in the present study (cf. ch. 4.1.3) can be described as a single-episode story with a recursive attempt section (cf. ch. 4.2.1 below).[37] Consequently, only the narrative components of a minimal story containing a single episode were described above, even though stories are generally far more complex. Nevertheless, even complex stories are variations of the basic one-episode story structure in that they may include several such episodes either following each other or embedded in the one-episode structure (see e.g. Stein & Glenn 1979, Mandel & Johnson 1977).

Based on the theoretical assumptions outlined so far, the coherence analysis conducted in the present study addresses the following questions:

- How coherent are participants' stories as measured by the number of narrative components?
- Are there any qualitative differences in coherence as measured by the frequency of individual components?

[36] Also REACTION (e.g. Stein & Policastro 1984).
[37] But cf. Bamberg & Marchman (1990, 1991, 1994) for a description as a multiple-episode story.

- Are there quantitative and/or qualitative differences in coherence attributable to grade, sex or L2 preschool experience?

2.4 Cohesion: From references to lexical cohesion

As explained in ch. 2.2, cohesion is a linguistic phenomenon. The development of the use of cohesive devices therefore reflects a linguistic development. To trace this linguistic development the present study tracks the development of the five "classical" categories of cohesion identified by Halliday and Hasan (1976), namely references, connectives, substitution, ellipsis, and lexical cohesion. Each one of these categories will be described in more detail in the following sections.

First of all, however, a general distinction has to be made according to the 'phoricity' of a cohesive device. That is, cohesive devices can refer either to elements in the extralinguistic ('exophora') or in the linguistic context ('endophora') (Hickmann 1991, Halliday & Hasan 1976). Even though the linguistic forms used for referring to linguistic and situational context are the same, exophoric reference does not directly contribute to cohesion (Halliday & Hasan 1976: 37).[38] The difference in phoricity thus corresponds to the developmental challenge children face: While even very small children are able to use cohesive devices exophorically (cf. ch. 3), they need to master the intralinguistic use of these cohesive devices, i.e. they need to learn "to use language *as its own context*" (Hickmann 1991: 157).

Additionally, endophoric cohesive relations can be either 'anaphoric', i.e. they point towards an antecedent in the preceding text (examples 2.3 & 2.4), or 'cataphoric', i.e. they point towards a referent in the following text (2.5).[39] However, cataphoric ties are comparatively rare and seldom cohesive (Halliday & Hasan 1976: 56).

2.3 *The cat ran away and it was never seen again.*

2.4 *The cat ran away and ∅ was never seen again.*

2.5 *It was gone, never to be found again. The cat had run away for good.*

[38] Most cohesive devices, i.e. substitution, ellipsis, connectives and lexical cohesion, are essentially endophoric relations, though, so that the distinction between endophora and exophora mainly becomes important for references (cf. Halliday & Hasan 1976).

[39] The term anaphora as it is used here refers to a relationship of co-reference, i.e. between one textual element and its antecedent (Crystal 2003: 24, Bamberg 1986: 231). In a more limited use of the term, it can be used to describe a relationship between a nominal expression and a pronoun (Bamberg 1986: 231).

The following questions will be investigated with regard to the overall use of cohesive devices:

- How cohesive are participants' stories as measured by the use of cohesive devices?
- Are there any qualitative differences as measured by the use of different types of cohesive devices?
- Are there quantitative and/or qualitative differences in cohesion attributable to grade, sex or L2 preschool experience?

Similar questions will need to be answered for each of the subtypes of cohesion under investigation. However, these additional questions will be presented in the corresponding sections below.

2.4.1 References

As all other cohesive devices, references refer to something else for their interpretation but "[i]n the case of reference the information to be retrieved is the […] identity of the particular thing or class of things that is being referred to" (Halliday & Hasan 1976: 31), i.e. the referent. Three types of references can be distinguished, namely personal, demonstrative and comparative references (ibid.: 31ff.). Personal references are established through pronominal markers referring to the identity of persons, objects, and events. They are expressed linguistically with the help of personal and possessive pronouns, e.g. s*he, mine, his.* Personal reference devices between clauses include relative pronouns, e.g. *who* or *which* (Quirk et al. 1985: 365). Demonstrative references, on the other hand, identify a referent by place or time on a temporal and spatial proximity scale through the use of demonstrative pronouns and adverbs, e.g. *this, here, now,*[40] as well as the definite article *the*. Comparative references, finally, are established by means of identity, similarity or difference expressed through adjectives[41] and adverbs such as *same, other, better, more* or *similarly*.

The analysis of the use of references in the present study addresses the following questions:

[40] Also *then* as a temporal adverb, e.g. *In my twenties I was pretty chubby. I loved cheeseburgers (back) then.* The demonstratives *then* and *now* have to be distinguished from the homonymous temporal connectives (cf. Halliday & Hasan 1976: 261).

[41] Halliday and Hasan's use of 'adjective' includes what other authors would rather label determiners and semi-determiners (e.g. *other*, cf. Biber et al. 1999: 280, Greenbaum 1996: 213).

- How often do participants use referential ties[42]?
- Are there differences in the use of references attributable to grade, sex or L2 preschool experience?

2.4.2 Substitution and ellipsis

Substitution and ellipsis are two closely related processes, which rely on structural relations for reference to a presupposed item (Halliday & Hasan 1976: 90), i.e. they act as a grammatical signal indicating that a presupposed item has to be recovered from the preceding text (ibid. 308). That is, substitution and ellipsis describe a relationship between linguistic items as opposed to reference, which describes a relationship between meanings (Halliday & Hasan 1976: 89).

Substitution refers to a presupposed element (a word or a group of words) by replacing it with a substitute item. Ellipsis, on the other hand, can be defined as a 'substitution by zero' (Halliday & Hasan 1976: 88ff.) or 'grammatical omission' (Quirk et al. 1985: 883) in that a sentence's grammatical structure contains an empty slot, which needs to be filled with a presupposed item from a neighboring part of the text (ibid.: 861, Halliday & Hasan 1976: 142ff.). Semantically, substitute item and ellipsis express general, "class" identity but at the same time non-identity of the actual referent – as opposed to reference, which presupposes the exact identity of referents (Halliday & Hasan 1976: 324). Additionally, substitution and ellipsis are often used to add a more specific contrast to the antecedent (ibid.: 314ff.).

Substitution and ellipsis each have three subtypes, namely nominal, verbal, and clausal substitution/-ellipsis. In nominal substitution the head of a nominal group[43] is substituted by *one*(*s*) or an entire nominal group by *same* – typically accompanied by *the* (Halliday & Hasan 1976: 91ff.; example 2.6).[44] In nominal ellipsis the head of a nominal group is omitted in the presupposing structure; its function can then be taken over by a modifier (example 2.7). Also, an entire nominal group can be omitted (example 2.8).

2.6 *I lost my <u>boots</u>. Now I've bought new <u>ones</u>.*

[42] In the following, the terms reference, connective, substitution etc. are used synonymously with referential ties, connective ties etc., since only cohesive devices forming ties were included in the analysis.
[43] Halliday's nominal phrase is defined as any constituent able to function as a subject or complement (Bloor & Bloor 1995: 259).
[44] A related process, namely replacement through a general word such as *thing* is covered by lexical cohesion in Halliday and Hasan's system, cf. ch. 2.4.4.

2.7 *I used to have three red blouses. Now I've got two ∅.*

2.8 *The little frog got out of his glass and ∅ ran away.*

Verbal substitution means substituting a lexical verb (and possibly additional elements) by a form of *do* (ibid.: 112ff.), while in verbal ellipsis the lexical verb ('lexical ellipsis'; example 2.9) or finite auxiliary and subject are omitted from the verbal group ('operator ellipsis'; example 2.10) and left to be inferred from the preceding text (ibid.: 170ff.).[45] In both types of verbal ellipsis additional elements can be left out.

2.9 *The frog may have escaped. Or he may not (have) ∅.*

2.10 *The boy had not found the frog and instead ∅ gone home.*

Clausal substitution is realized by *so* and *not* presupposing an entire clause as in example (2.11) (ibid.: 130ff.). In clausal ellipsis, on the other hand, either subject plus finite element of the verbal group ('modal ellipsis'; example 2.12) or the remainder of the verbal group as well as any complements and adjuncts ('propositional ellipsis'; example 2.13) are omitted. Modal ellipsis typically includes operator ellipsis in the verbal group (cf. example 2.10), propositional ellipsis includes lexical ellipsis (cf. example 2.9) (ibid.: 199). In other types of clausal ellipsis an entire clause is left out, e.g. in yes/no ellipsis (example 2.14). All forms of clausal ellipsis typically occur in question-answer sequences (cf. ibid.: 196ff.).

2.11 *Mary will drive tomorrow. If not, I will.*

2.12 A: *What are you doing?* B: *∅ reading a book.*

2.13 *You can come to my house tomorrow, if you want ∅.*

2.14 A: *Have you done your homework?* B: *Yes ∅.*

The analysis of the use of substitution and ellipsis in the present study addresses the following questions:

- How often do participants use substitution/ ellipsis?
- Are there differences in the use of substitution/ ellipsis attributable to grade, sex or L2 preschool experience?

2.4.3 Connectives

Connectives make explicit the semantic relationship between one clause or sentence and another and thereby relate text sequences which are structurally unconnected

[45] For a definition of the Hallidayan verbal group cf. Bloor & Bloor 1995.

(Halliday & Hasan 1976: 226ff.).[46] Formally, this can be accomplished with the help of conjunctions, e.g. *and*, simple or compound adverbs, e.g. *then*, *finally* or *afterwards*, prepositional phrases with or without *that*, e.g. *as a result* (*of that*) or *on the contrary*, and other complex connectives, e.g. *one day later*.

Semantically, most connectives express one of four basic relationships, namely additive, temporal, causal and adversative (e.g. Biber et al. 1999: 79ff., Quirk et al. 1985: 915ff., Halliday & Hasan 1976: 226ff.).[47] Additive connectives make explicit the underlying structural relationship of coordination, which joins linguistic elements to function as one complex element of structure in the sense of one element being an addition to the other. Temporal connectives express a relationship of sequence or simultaneity, while causal connectives mark a relationship of causality and adversative connectives express a contrast. There is no simple correspondence between the form and meaning of a connective, however. The coordinating conjunction *and*, for example, can express several types of relations besides additive, e.g. temporal (*I bought some milk and went home*) or adversative (*I asked him to help me and he did not want to*) (cf. Quirk et al. 1985: 930ff.)

In addition to formal and semantic criteria, connectives can be distinguished according to the syntactic role of the elements they link, i.e. coordination in a paratactic or subordination in a hypotactic construction. *But*, for example, expresses an adversative relation between two coordinate clauses, while *because* marks a causal relationship between a main clause and the subordinate clause it introduces.

In studying the use of connectives it is crucial to keep in mind that the semantic relations between clauses and sentences do not have to be made explicit (e.g. de Beaugrande & Dressler 1981, Mandler & Johnson 1977, Halliday & Hasan 1976: 229).[48] Two sequentially stated event clauses, for example, are assumed to reflect the order of occurrence of these events, i.e. listeners automatically infer that the event stated first also occurred first; an explicit linguistic marking of this relationship is therefore only necessary if the sequence is out of order (e.g. Hickmann 2003: 95,

[46] That is, text sequences which are not connected by being part of the same grammatical construction (Halliday & Hasan 1976: 6f.).

[47] No general agreement exists on the number of categories and subcategories. Other authors recognize additional categories, for example concessive and conditional (e.g. Biber et al. 1999, Quirk et al. 1985), which Halliday and Hasan would subsume under the category adversative and causal, respectively. Halliday and Hasan, on the other hand, recognize an additional "residual" category 'continuatives' (1976: 267ff.).

[48] See also the distinction Halliday and Hasan make between 'external connectives', which make explicit relations between events, processes or states, and 'internal connectives', which are language-internal structuring devices (1976: 239ff.).

Peterson & McCabe 1991: 31). Concerning the development of the use of connectives this means that a lack of connectives does not automatically relate to a lack of concept. Even children at age 3, for example, are able to attribute actions to goals of the person acting (Sodian 2002: 457). Nevertheless, children at that age rarely use connectives at all and much less explicitly causal ones (cf. ch. 3).

The analysis of the use of connectives in the present study addresses the following questions:

- How often do participants use connective ties?
- Are there differences in the use of connective ties attributable to grade, sex or L2 preschool experience?

2.4.4 Lexical cohesion

Lexical cohesion is the most frequent type of cohesion in texts and thus also carries most of their cohesive load (e.g. Bae 2001, Hoey 1996, Hasan 1984). It can be defined as "the cohesive effect achieved by the selection of vocabulary" (Halliday & Hasan 1976: 274), i.e. ties created through the lexical (rather than textual) relations between nouns, adjectives, lexical verbs and, to a lesser extent, also adverbs (e.g. Beigman Klebanov & Shamir 2005, Hoey 1996: 6ff.). Such relations between lexical items can involve a general semantic relationship, e.g. a hyponymic relation as between *frog* and *animal*, or an 'instantial' one, i.e. a relationship created by the content of the specific text (Hasan 1984: 202, Halliday & Hasan 1976: 289). Such an instantial lexical relationship is shown in example (2.15), where *dog* and *Bello* denote the same referent just as *frog* and *Bingo* denote the same referent.

2.15 1[One night, Tim # had a dog, 2[names Bello,] and a frog # 3[who called # Bingo.] # 4[they were very happy] 5[that they had b/ Bingo.] (C5-G4-3)

Two large subcategories of lexical cohesion can be distinguished: 'Reiteration' and 'collocation' (Halliday & Hasan 1976: 274ff.). Reiteration refers, first of all, to ties created by repetition, which range from simple reproductions, e.g. *boy – boy*, to the 'complex repetition' of lexical items that share a lexical morpheme but are not formally identical, e.g. *boy – boyish* (Halliday & Hasan 1997 [1985]: 81, Hoey 1996: 55). Reiteration also includes ties by synonymy and near-synonymy, recurrence to a hyperonym or the use of a 'general noun' such as *thing* or *man* (Renkema 2004: 105, Hallliday & Hasan 1976: 277ff.). At the same time the relationship between two

lexical items tied by reiteration can be described on a continuum ranging from mere identity of form (homonymy) to identity of referent (Halliday & Hasan 1976: 283). Non-identity of referent is usually compensated for by a higher number of additional lexical ties as compared to instances with identical reference (ibid.).

Collocation refers to the cohesive effect between lexical items that "are in some way associated with each other in the language" (Halliday & Hasan 1976: 285), i.e. lexical items which regularly co-occur. This ranges from being systematically connected through semantic relations such as co-hyponymy, (co-)meronymy and opposition (e.g. antonymy, converses), i.e. those relations not included in the reiteration category, to habitual co-occurrence with no readily specifiable sense relation, e.g. *laugh – joke* or *cut – knife* (ibid.). Collocation in this latter sense of co-occurrence may in most cases be attributable to relatedness through schema membership (cf. ch. 2.1), e.g. a TELLING A JOKE and a CUTTING schema in case of the examples given. The strength of the cohesive effect achieved by collocation is determined by the lexical items' proximity in the lexical system, their distance in the text at hand, and their overall frequency in the language system (ibid.: 290).

The analysis of the use of lexical cohesion in the present study addresses the following questions:

- How often do participants use lexical cohesion?
- Are there differences in the use of lexical cohesion attributable to grade, sex or L2 preschool experience?

2.5 Narrative production: Summary and research questions

As explained in the previous sections, two fundamental dimensions of (narrative) discourse can be distinguished: Coherence and cohesion. Cohesive devices are text-based, linguistic devices, so that the development of their use reflects a general development in linguistic abilities. Coherence, on the other hand, which in most general terms refers to the content structure of a text, is based on narrative schemata. The development of narrative coherence therefore reflects a development in cognitive abilities. By exploring the coherence and cohesion of IM students' L2 narratives I will thus trace their cognitive and linguistic development from first to

fourth grade.[49] Additionally, a possible relationship between (the development of) coherence and cohesion will be explored. That is, two main research questions are addressed in the present study:[50]

1) Is there a cognitive development from grade 1 to 4 as measured by L2 narrative coherence?
2) Is there a linguistic development from grade 1 to 4 as measured by L2 narrative cohesion?

Additionally, the following question will be addressed:

3) Is there a relationship between (the development of) L2 cohesion and coherence?

[49] It may not immediately be evident how L2 data could reflect cognitive development. However, the overview of previous research given in ch. 3 will show that children's narrative organization evolves very strongly within the age range of the present study's participants. Additionally, it will be shown that this development is evident not only in an L1 but also in an L2 once a certain linguistic threshold has been passed.

[50] The research questions presented in the preceding sections should be seen as outlining necessary steps towards answering these main questions.

3 The development of storytelling

Compared to well-researched areas in (second) language acquisition, such as morpho-syntax or semantics, few studies have been conducted on narrative coherence and cohesion and their development. At the same time the majority of these latter studies investigated L1 speakers. That is, the field of narrative coherence and cohesion is strongly underresearched not only for L1 but even more so for L2 and bilingual data. In order to obtain as complete a picture as possible of the development of coherence and cohesion the present overview will include findings on L1 as well as L2 and bilingual speakers; this is especially important since participants of the present study are child L2 learners, whose L1 development has to be considered far from concluded.

Sex and preschool experience will not be taken into account for the following reasons: Even though sex is generally included among the potential factors of influence on (second) language acquisition, its impact is far from clear (cf. the overview in Ellis 1994). The influence of sex on cohesion and coherence has rarely been investigated and if sex was considered, it was typically found of no influence to overall coherence and cohesion (e.g. Allen et al. 1994, Stenning & Mitchell 1985, Peterson & McCabe 1983, Botvin & Sutton-Smith 1977). However, some studies found at least minor qualitative differences attributable to sex (e.g. Stenning & Mitchell 1985, Peterson & McCabe 1983, cf. also the overview in Hendrickson & Shapiro 2001). With respect to preschool experience, there is a severe lack of previous studies. As already mentioned, the majority of studies on coherence and cohesion has focused on monolingual speakers' discourse and far fewer studies were conducted on L2 learners (especially young ones) and/or bilinguals. To the best of my knowledge, no study has so far compared young L2 learners with and without L2 preschool experience, so that no prior findings are available which could be included in the overview.

Additional influence factors, for example comprehension skills (Cain 2003) or mothers' encouragement in mother-child storytelling activities (Kang et al. 2009), will also not be covered in the developmental overview, since they are not investigated in the present study. It should be kept in mind for the interpretation of results, however, that variables such as these may also have been found to influence results had they been investigated.

In the following sections I will give an overview of previous findings on the development of narrative coherence in L1 learners' discourse (ch. 3.1) as well as in L2 and bilingual learners' discourse (ch. 3.2). Then I will summarize findings on the development of cohesion in L1 learners' discourse (ch. 3.3) as well as in L2 and bilingual learners' discourse (ch. 3.4). It will become clear that – due to the age-range under investigation – a development in the coherence and cohesion of participants' discourse is to be expected in the present study, irrespective of whether L2 or L1 data is analyzed. After the developmental overview of coherence and cohesion, an overview of findings on the relationship between these two discourse characteristics will be given (ch. 3.5). This latter overview will show that the results of earlier studies show no consistent pattern. The chapter concludes with a summary in which several hypotheses are formulated with respect to the research questions raised in the previous chapter.

3.1 The development of narrative coherence in L1 acquisition

Children's earliest narrative productions seem to be accounts of personal experiences ('personal narratives') in reaction to adult questions. Thus, at age 2 children are already able to relate personal narratives if heavily prompted and encouraged by their parents, i.e. "storytelling" rather consists in a joint activity of parental questions and child answers (McCabe & Peterson 1991a: 250).

Between age 3 and 4 children's narrative productions remain short strings of events without any global organization structure, which are – in spontaneous production – most often either again accounts of personal experiences or general event representations, i.e. scripts (Hudson & Shapiro 1991: 121ff., McCabe & Peterson 1991b: 246f., Nelson & Gruendel 1986: 41ff., Seidman et al. 1986: 167ff.).

With respect to fictional stories it was found that at age 3 children are already able to respond to a picture-elicited storytelling task, even if they still have difficulties to

concentrate and need a very high degree of interviewer support (Berman & Slobin 1994b: 58ff.). The resulting story productions consist of a very limited number of narrative components and mainly correspond to descriptions of the event contents of some of the pictures, i.e. two or three events connected by *and then* (e.g. Berman & Slobin 1994b: 46, Berman 1988, Peterson & McCabe 1983).

While there is no evidence for an underlying story schema in production at that age, this does not, however, exclude the possibility that children are using a story schema for comprehension and recall (Trabasso et al. 1995, Seidman et al. 1986; cf. also the overview in Hudson & Shapiro 1991: 100f.).

From age 3 to 4 onward children's personal narratives and fantasy stories gradually develop from being descriptions of isolated objects, events or states to event sequences. Especially from age 5 on there is also increasing evidence for some children to arrange their stories hierarchically, i.e. in relation to an overall goal-plan, even if such a global organization remains the exception (Berman & Slobin 1994, Hudson & Shapiro 1991, Berman 1988, Seidman et al. 1986: 167f., Peterson & McCabe 1983).

By age 8 to 9 the development of scripts and personal narratives seems to be relatively complete (Hudson & Shapiro 1991: 123f., Peterson & McCabe 1983). Children's ability to tell a globally organized fictional story, on the other hand, continues to develop.

Around age 9 to 10 a large number of children's fictional stories finally start to resemble those of adults, whose stories are organized in relation to a global narrative structure reflecting the use of a narrative schema (e.g. Berman & Slobin 1994b: 58, Berman 1988). This increase in stories with a global organization is reflected by an increase in the overall number of narrative components as well as in the frequency of the individual components (e.g. Reilly et al. 1998, Berman & Slobin 1994b: 53f., Hudson & Shapiro 1991, Berman 1988).

Some narrative components seem to be easier to acquire than others, however, in the sense that they are produced from an earlier age on. At the same time research results for individual components are not easy to compare, due to differing definitions and approaches to narrative structure (but cf. the developmental overview in Hudson & Shapiro 1991: 89ff.): Berman and Slobin and their colleagues (1994), for example, investigated (among other components) the development of the 'onset', which in their study corresponds to one part of what was outlined as story grammar's

initiating event in ch. 2.3.2, in the story productions of children age 3 to 9 and adults. They found that across five languages the large majority of children (78%) had acquired the onset by age 5. Hudson and Shapiro (1991), on the other hand, found that their 'problem' component, which corresponds to story grammar's initiating event, was realized by only around 34% of first graders (mean age 6;7). A percentage of realizations comparable to Berman and Slobin, namely over 70%, was not reached before third grade (mean age 8;7), i.e. almost four years later. Previous studies only seem to fully agree that by age 4 to 5 children include at least a primitive setting or 'orientation' component, whose realization and complexity increases with age (e.g. Berman 2001, Peterson & McCabe 1983, Pradl 1979).

As opposed to children, adults' narrative productions all have a globally organized structure (Lanza 2001, Berman & Slobin 1994b: 75). Additionally, they are much more elaborate and detailed, e.g. in references to inner states (Berman & Slobin 1994b: 80). Further developments into adult age include the "individualization" of stories (Berman & Slobin 1994, Berman 1988). As Berman and Slobin (1994b: 74f.) note, narrative productions at age 9 (i.e. their oldest group of child participants) can best be described as stereotypical because there is so little individual variation. Adult stories, on the other hand, are very heterogeneous.[1]

To sum up, children's storytelling abilities evolve continuously from script-like personal narratives and picture descriptions (up to ca. age 3) to event sequences (from ca. age 3 on) and finally to globally organized narratives (from ca. age 5 on). While the structure of scripts, i.e. general event representations, and personal narratives can be considered acquired by around age 8, the ability to tell conventionally structured fantasy stories is a relatively late development. That is, the use of a global organization structure (story schema) can be found in the majority of children's stories around age 9 but it continues to develop into adulthood – accompanied by a greater individualization of storytelling styles.

If children's data is regarded in its own right, it can thus be argued that children successively acquire different types of story schemas. That is, very young children have acquired a "joint activity" schema, which corresponds to the way in which caregivers seem to engage in a storytelling task (Berman & Slobin 1994b: 60,

[1] Children's more stereotypical view of stories is also reflected in their story goodness ratings: Stein and Policastro (1984: 151), for example, found that adults (elementary school teachers) more readily included non-prototypical passages, e.g. non-goal-based reactive sequences, into the story category than did 7- to 8-year-olds.

Trabasso & Rodkin 1994, McCabe & Peterson 1991, Shapiro & Hudson 1991). Preschool and early school children, on the other hand, have acquired a descriptive model of storytelling (Stenning & Mitchell 1985. Cook-Gumperz & Green 1984). Older children, finally, start to acquire an adult-like schema, which involves a global organization structure as described, for example, by story grammars.

3.2 The development of narrative coherence in L2 and bilingual acquisition

The results obtained in studies focusing on the narrative coherence of child L2 learners and/or bilinguals confirm the age-related use and development found in studies on monolinguals. That is, the development of a global narrative structure begins at around age 5 and the number of narrative components as well as the frequency of the individual components increase with age (e.g. Montanari 2004, Akinçi et al. 2001, Kupersmitt & Berman 2001, Severing & Verhoeven 2001).[2]

Whether narrative training is received in an L1 or L2, e.g. due to being enrolled in an immersion program, does not seem to have adverse effects on the development of storytelling abilities (Almgren et al. 2008 & 2007, Hüttner & Rieder-Bünemann 2007, Laurén 1998). Almgren et al. (2007), for example, found that L1 Spanish students in a Basque immersion program, who had been trained in storytelling via their L2 Basque, were able to transfer this knowledge to narrative production in their L1 Spanish.

Several studies have shown, however, that a certain linguistic threshold has to be passed to ensure that once the cognitive schema is available so are the linguistic means to express it in the L2 or, in the case of very early bilinguals, in both languages (e.g. Almgren et al. 2008 & 2007, Hüttner & Rieder-Bünemann 2007, Montanari 2004, Severing & Verhoeven 2001). Severing and Verhoeven (2001), for example, investigated the narrative productions of L1 Papiamento speakers educated in Dutch (mean age 5;2 to 10;4). On the one hand they found a consistently higher performance in the L1, but on the other hand they also found that after four years of schooling the differences between participants' L1 and L2 story productions had started to become smaller. This trend towards a similar performance in the L1 and L2 is supported by the results of studies on very early bilinguals, which observed

[2] However, most studies on narrative development in an L2 focused on participants older than age 5 so that there is a severe lack of more studies on young L2 learners just as for early bilinguals.

almost no difference in coherence between bilinguals' two languages once the necessary linguistic means had been acquired (e.g. Akinçi et al. 2001, Kupersmitt & Berman 2001, Lanza 2001, Viberg 2001).

The impact of a linguistic threshold is even more evident in older L2 learners, who can be expected to have full command of a story schema as well as – from their L1 experience – a general understanding of how to express such structures linguistically. Myles (2003), for example, studied the narrative development of British L2 French learners after one (8th grade) and two years of classroom instruction (9th grade). She found that not only the number of propositions, but also the frequency of individual narrative components increased as a function of the time of exposure, even though her eighth and ninth graders should already have had a well-developed narrative schema available for use. A comparison of Myles' (2003) results with the studies on young L1 and L2 learners also indicates that time of exposure (as well as related factors such as the amount of input) becomes the crucial factor for older learners, while in younger learners an overlap of age-related cognitive and linguistic development with an exposure-related L2 development needs to be expected.

Further confirmation for the impact of linguistic proficiency on coherence in L2 narratives comes from studies conducted on immersion students, who receive a much higher amount of language input than L2 learners in regular programs. Immersion students have been found, first of all, to produce more coherent stories than students enrolled in regular foreign language teaching (Hüttner & Rieder-Bünemann 2007). At the same time IM students have been found to perform quite similarly to monolingual comparison groups. Laurén (1998), for example, studied L2 Swedish stories produced by Finnish fifth graders, who had been enrolled in a Swedish immersion program for five years, preceded by one year in an immersion kindergarten. Laurén found that the IM group's mean number of components was only slightly lower than the one of the L1 Swedish comparison group (Laurén 1998), i.e. both L1 and L2 speakers produced comparably coherent stories.

To sum up, the development of coherence in L2 learners and bilinguals parallels the development found in L1 acquisition, i.e. an increase is to be expected in the overall number of components as well as in the frequency of individual components. Once a narrative schema has been acquired, it can be used for narrative productions in any language. However, L2 learners also need to have acquired the necessary linguistic means to express the schema in the L2.

3.3 The development of cohesion in L1 acquisition

Children have been found to use cohesive devices in conversations and personal narratives from age 2 on and in fictional stories at age 3 (Berman & Slobin 1994b: 67, Peterson & Dodsworth 1991; cf. also the overviews in Shapiro & Hudson 1991 and Bamberg 1987). However, few studies on L1 and also L2 narratives have investigated all five types of cohesive devices differentiated by Halliday and Hasan (1976) – most studies focused instead on character references and/or connectives – and even fewer have done so for oral data. The overview of cohesion in L1 narrative productions given in the present section and the subsequent overview of cohesion in L2 narratives will therefore include findings from studies not only on oral but also on written data. Since oral and written stories are not necessarily directly comparable, however, it will be indicated in the text where studies obtained their results from written data.

A notable exception to studies covering only references and/or connectives are Peterson and Dodsworth (1991), who studied the development of cohesion in two-year-olds' personal narratives over a span of one and a half years. Peterson and Dodsworth found that even at age 2 children were able to use most of the subcategories differentiated by Halliday and Hasan (1976); exceptions were comparative references, nominal ellipses, and substitutions. By age 3 their participants were able to use all of the cohesive devices distinguished by Halliday and Hasan. Peterson and Dodsworth also found that the overall number of cohesive ties increased significantly over the duration of their study, which was mainly attributable to a rise in the use of personal references and conjunctions. The (sub-) categories of ellipsis and substitution, on the other hand, decreased or remained fairly stable.

Additionally, Peterson and Dodsworth found a stable order of frequency for the five types of cohesion distinguished by Halliday and Hasan: Across all narrative productions lexical ties were used most frequently, followed by references and conjunctions. Ellipses and especially substitutions, on the other hand, were used very infrequently. This order of frequency was confirmed in studies on the written L1 narratives of children (e.g. Cameron et al. 1995, Spiegel & Fitzgerald 1990, Crowhurst 1987, Yde & Spoelders 1985) as well as, for example, in adults' essay writing (Neuner 1987).

Research on written narratives has shown that by around age 8 to 9 (at the latest) a ceiling is reached in the number of cohesive devices: Studies comparing 8- to 9-year-olds with older peers found no further significant increase in cohesiveness or even a decrease in some of the subcategories (e.g. Manhardt & Rescorla 2002, Spiegel & Fitzgerald 1990; but cf. Yde & Spoelders 1985). Instead, a qualitative shift in the subcategories of cohesive devices was found for older children, for example a decrease of repetitions as opposed to an increase of synonyms and collocations within the lexical cohesion category (Crowhurst 1987). In oral narratives the same phenomenon has been shown for connectives (e.g. Berman & Slobin 1994d: 609f., Berman 1988; cf. also below).

Most prior studies on cohesion have, as already mentioned, focused on connectives and references. In the following I will therefore also focus on research results for these two cohesive devices. With respect to connectives it was found that, although 2- and 3-year-olds already use connectivity markers in their narratives, their number is comparatively low (e.g. Berman & Slobin 1994, Berman 1988). In the following years, i.e. up to around age 9, children's use of connectives increases (e.g. Berman & Slobin 1994c: 177ff., Strömqvist & Day 1993, Hudson & Shapiro 1991, Bennett-Kastor 1986). Berman and Slobin (1994c: 175ff.), for example, found that at age 3 their monolingual English-speaking participants marked under 50% of the relations between clauses, while 5- and 9-year-olds did so in over 85% of clauses. However, around age 10 (at the latest) the number of connectives seems to decrease again (Berman & Slobin 1994d: 609f., Berman 1988).

With age the use of connectives also becomes more sophisticated. This means, first of all, that children's use of connectives shifts from a pragmatic to a discourse function and, secondly, that the variety and complexity of connectives increases – be it in oral (e.g. Verhoeven et al. 2002: 148ff., Berman & Slobin 1994c: 175ff., Strömqvist & Day 1993, Hudson & Shapiro 1991) or written narratives (e.g. Verhoeven et al. 2002: 148ff., Crowhurst 1987, Fitzgerald & Spiegel 1986, McCutchen & Perfetti 1982: 128). Thus, 3-year-olds typically produce utterance-initial deictics (e.g. *here*) indicative of a picture description mode or they use the coordinator *and* with a discourse-pragmatic function, i.e. to indicate that "more is to come" (Berman & Slobin 1994c: 175f., Berman 1988). Nevertheless, even very young children are able to produce structurally subordinating and semantically temporal, causal or adversative connectives, especially if such relationships are

encouraged by the elicitation material (Shapiro & Hudson 1997, Berman & Slobin 1994c: 176).

In the following years children's preference gradually shifts to temporal expressions of sequentiality, especially *and then* (Berman & Slobin 1994c: 177ff.). By age 9 children still rely heavily on this temporal sequencing but they also produce more advanced connectives, e.g. connectives expressing causal and adversative relations or syntactically complex ones. Adults, finally, use explicit expressions for temporal sequence (such as *and then*) only marginally, since they assume temporal sequence to be the default relation between clauses (Berman & Slobin 1994d: 610, Peterson & McCabe 1991: 47). Instead, adults produce an even larger variety of connective devices, e.g. more subordinators, and they integrate connectives more flexibly into the sentence structure, e.g. by using connectives such as *finally* also in non-clause-initial position (Verhoeven et al. 2002: 149f., Berman & Slobin 1994c: 180ff.).

These maturational changes in the use of connectives are indicative for a similar development as the one found for narrative structure. That is, children's use of connectives reflects a gradual movement away from a simple picture-by-picture description towards, first of all, a temporal structuring of events and, finally, a thematically-motivated global structure (Aksu-Koç & von Stutterheim 1994: 419, Berman & Slobin 1994c: 175, Strömqvist & Day 1993).

Not only the use of connectives but also that of references becomes more sophisticated with age. This has been shown especially for personal references, where the changes reflect the developments found for connectives, i.e. from (a) a deictic to an intralinguistic use of referential devices and (b) towards global organization strategies (e.g. Hickmann 2003 & 1991, Shapiro & Hudson 1997 & 1991, Wigglesworth 1990, Bamberg 1987 & 1986, Karmiloff-Smith 1985).

3-year-olds use referential devices mainly deictically but children's references gradually become intralinguistic (Hickmann 2003 & 1991, O'Neill & Holmes 2002, Karmiloff-Smith 1985). Once referential devices are used intralinguistically, children pass through several stages characterized by different referential strategies (e.g. Hickmann 2003 & 1991, Shapiro & Hudson 1997 & 1991, Bamberg 1987 & 1986, Karmiloff-Smith 1985): The intralinguistic use of referential devices is first of all dominated by the immediate linguistic context, i.e. children use a locally-motivated strategy (predominant from ca. age 3 to 5). This evolves into strict adherence to a

'thematic subject strategy' (Karmiloff-Smith 1985), i.e. a globally-motivated use of referential devices (predominant from ca. age 5 to 7), and then into an adult-like use, which integrates local and global strategies (from ca. age 8 on).

Evidence on the respective age ranges is controversial, though. While Karmiloff-Smith (1985), for example, found that the thematic subject strategy develops relatively late, other studies report that around age 4 and 5 the thematic subject strategy is already most common (Shapiro & Hudson 1991, Bamberg 1987 & 1986). These differences in results could, however, be attributable to differences in task condition, task material or language typology (cf. Shapiro & Hudson 1997 & 1991, Bamberg 1986, and the discussion in Hickmann 1996: 205ff.). There is general agreement at least that children's reference system continues to develop even beyond age 8 (e.g. Hickmann 2003 & 1991, Shapiro & Hudson 1997: 40, Wigglesworth 1990, Karmiloff-Smith 1985).

To sum up, some cohesive devices have been observed in children's narratives from age 2 on. The number and type of cohesive ties increases with age and at the same time the use of cohesive devices becomes more sophisticated. However, evidence especially from written narratives indicates a ceiling in the number of cohesive ties, which seems to be reached around age 8 to 9. After that the use of some subcategories can even decrease and further development seems to be limited to qualitative changes within the subcategories.

3.4 The development of cohesion in L2 and bilingual acquisition

Studies conducted on the use of cohesive devices by L2 or bilingual subjects parallel the results obtained in studies on monolinguals with respect to the overall order of frequency and the general development of the use of cohesive devices. That is, first of all, the order of frequency of the categories distinguished by Halliday and Hasan (1976), namely lexical ties, references, connectives, substitution and ellipsis, was found to be the same – be it in studies on child L2 learners (Bae 2001) or adult L2 speakers (Norment 1995, but cf. Kang 2005). Bae (2001), for example, investigated the use of all five categories of cohesion in the written L2 stories of first and second graders enrolled in immersion and English-only programs. Bae found, first of all, that the children were able to use all five types of cohesive devices. Just as in studies on L1 speakers, however, participants most frequently used lexical ties, which

accounted for 56% of their cohesive devices, followed by references (32%) and (coordinating) conjunctions (12%). The use of ellipses and substitutions was again found to be marginal (both below 1%).

With respect to the development of the use of cohesive devices, it was found that the number and variety of connectives increases with age also in an L2 (Bae 2001, Severing & Verhoeven 2001), respectively in both languages in the case of very early bilinguals (Montanari 2004, Aarsen et al. 2001), just as L2 and bilingual speakers' reference strategies become more sophisticated with age (Montanari 2004, Severing & Verhoeven 2001). In older L2 learners, similar developments take place (Myles 2003, Norment 1995), which are again rather attributable to factors such as exposure time and/or input (and the corresponding proficiency level), however, while in young learners' L2 productions these factors seem to overlap with maturational effects (cf. also ch. 3.2).

Even if L2 learners' general development in the use of cohesive devices parallels the one of L1 learners, older L2 learners have typically been found to perform differently from monolingual speakers of the respective target language both in oral and written narratives (e.g. Kang 2005 & 2004, Myles 2003, Reid 1992) L2 speakers/writers have been found, for example, to overuse coordinate conjunctions (Reid 1992) and nominal referential devices (Kang 2004). However, learners' use of L2 cohesive devices has also been found to differ from their L1 productions (e.g. Kang 2005, Norment 1995). Thus, the differences between L2 learners' use of cohesive devices and the one of monolingual controls seem to be attributable an intermediary stage in their interlanguage.

The results obtained by Laurén (1998: 528) point in the same direction. Laurén analyzed the use of anaphoric references by Finnish children attending the fifth year of a Swedish immersion program, which had been preceded by another year in an immersion kindergarten. She found that almost all immersion students used the same linguistic means in Swedish as the L1 comparison group, even though these linguistic means are different from those available in Finnish.

To sum up, the cohesion categories' general order of frequency and the development of cohesion in young L2 learners' and bilinguals' narratives parallel the ones found in L1 acquisition: The number and variety of ties increases with age and their use becomes more sophisticated. In the case of older learners the same developmental process has been found with time of exposure and not age as the

determining factor. At the same time L2 learners perform differently from L1 speakers, while they are still reorganizing their interlanguage system.

3.5 The relationship between coherence and cohesion

As the developmental overview of coherence and cohesion has shown, a number of studies has investigated the development of these two discourse characteristics in children's narrative productions. Far fewer L1- and especially L2-studies, however, have explored the relationship between the coherence and cohesion of children's texts and these studies have had conflicting results (cf. also chapter 2). Additionally, a comparison of findings is made difficult by the varying operationalizations of both coherence, e.g. a discourse-based measure such as story grammar as opposed to external coherence ratings, and cohesion, e.g. the number of ties as opposed to reference strategies.

Studies using two discourse-based measures typically found a relationship between coherence and at least some cohesion measures, usually references and connectives (e.g. Cain 2003, Shapiro & Hudson 1997 & 1991, Spiegel & Fitzgerald 1990, Stenning & Mitchell 1985). This latter finding may, however, be biased somewhat by the general preference for investigating connectives and references, which was discussed earlier in this chapter.

Shapiro and Hudson (1997, 1991), for example, investigated the development of coherence and cohesion in the oral narratives of preschoolers (mean age 4;8) and first graders (mean age 6;8). They operationalized coherence with the help of an overall complexity measure for narrative structure (based on the number and types of episodic components) and cohesion with the help of a subordination index, the subtypes of connectives and participants' reference strategies. Shapiro and Hudson found that, regardless of age, children producing structurally more complex narratives also used more complex language (i.e. subordination) and more sophisticated cohesive devices. Similarly, Cain (2003), who studied coherence (operationalized as completeness of the event structure) and connectives in oral stories produced by 6- to 8-year-olds, found a positive correlation between story coherence and the use of sophisticated connectives in her participants' narratives.

However, not all studies' findings unambiguously confirm this pattern. Stenning and Mitchell (1985), for example, studied the oral story productions of children age 5 to 10. They found that children using sophisticated connectives had a tendency to

score better for story content and answers to explanatory questions. At the same time Stenning and Mitchell also found a strong interindividual variation, however, i.e. many of their child participants who told sophisticated stories used unsophisticated connectives. Additionally, reference strategies were not found to be indicative of the two coherence measures in Stenning and Mitchell's study.

Several studies, especially on written narrative productions, operationalized coherence with the help of rating scales, which were administered by teachers, for example, and compared the coherence scores to the use of cohesive ties. Fitzgerald and Spiegel (1986), for example, investigated the relationship between coherence and cohesion in third and sixth graders' written stories. Across grades, they found a trend towards a negative correlation between coherence ranking and the number of cohesive ties, i.e. stories with fewer ties were given higher coherence rankings.

To the best of my knowledge, only one study on L2 learners has explored the relationship between coherence and cohesion, namely Bae (2001), and this study also used a rating scale approach. Bae investigated narrative texts written by first and second grade L2 English learners and found that the number of at least some cohesive ties was related to coherence scores. That is, the number of both lexical and referential ties was a significant predictor for coherence ratings. Connectives, ellipses and substitution, on the other hand, had no predictive value. The difference between Bae's and Fitzgerald and Spiegel's results could of course be attributable to differences between L1 and L2 speakers. However, a much larger number of comparable studies is needed to make any claims about differences and/or similarities between L1 and L2 studies – especially when keeping in mind that, as described earlier, even largely comparable L1 studies have yielded somewhat contradictory results.

To sum up, this overview has shown that even studies with comparable approaches do not yield consistent results with respect to the relationship between coherence and cohesion in children's narrative productions. However, L1 studies show at least a trend towards a correlation between coherence and the use of sophisticated connectives. The only prior research conducted on L2 data, which used a different approach from the one in the present study, found a relationship between lexical and referential ties and coherence rating.

3.6 Summary

As this review of earlier research has shown, children's L1 narrative productions become more coherent with age. That is, they are increasingly built around a global problem-resolution structure (top-down organization), which is indicative of an underlying narrative schema, instead of being a simple chaining of events (bottom-up organization). This development seems to be most prominent up to the age of 9, but it continues beyond that age.

With respect to the use and development of cohesive ties in the L1, this overview has shown that even 3-year-olds are able to produce almost all categories of cohesive devices distinguished by Halliday and Hasan (1976). With age the number of cohesive devices increases and their use becomes more sophisticated in a way that reflects the development from local to global organization found for narrative coherence. However, around age 8 to 9 a ceiling in the number of cohesive devices may be reached after which the development is mainly limited to qualitative changes within the individual categories.

The overview also presented evidence for a general order of frequency, i.e. lexical ties make the strongest contribution to cohesion, followed by reference and connectives, while ellipsis and substitution contribute only marginally.

This review of earlier studies has also shown that the coherence of L2 productions and its development do not differ from those evident in L1 narratives once the speaker has acquired the necessary linguistic means in the L2. Instead, narrative schemata as production plans seem to be transferable from one language to the other, at least within the same cultural context. With respect to cohesion in L2 learners' narratives previous studies showed, first of all, that the same order of frequency holds true. Secondly, the general development of cohesion parallels the one of L1 learners, i.e. a maturational and/or exposure-related increase in the number and sophistication of cohesive devices is to be expected. At the same time L2 learners' use of cohesive devices typically differs not only from L1 comparison groups (at least before full native-like control has been obtained) but also from learners' use in their L1.

Regarding the relationship between coherence and cohesion in narrative productions it was shown that no clear picture emerges for either L1 or L2 discourse. Research on L1 data found a tendency, however, towards a correlation between at least some cohesion measures and coherence. The only study on L2 data found a

relationship between coherence ratings and the number of both lexical and referential ties.

Based on the developmental overview given in the previous sections the following basic conclusions can be drawn for the design of the present study:

1) The participants' age range (mean age 6;7 to 9;7) is suitable to study a development in narrative coherence and cohesion
2) This development in narrative coherence and cohesion can also be traced in L2 data (keeping in mind a potential linguistic threshold).

With respect to the research questions presented in the previous chapter several hypotheses can now be formulated. First of all, regarding the impact of the three learner variables under investigation, it is assumed that

(1) Sex does not have a significant influence on the narrative coherence or cohesion of participants' stories
(2) Grade/age has a significant influence on the narrative coherence and cohesion of participants' stories.

As indicated at the beginning of the chapter, no previous studies are available on the influence of L2 preschool experience on coherence and cohesion, so no hypothesis will be put forward.

Regarding a grade/age-related development, several additional hypotheses can be formulated:

(3) Participants' stories become more coherent from first to fourth grade as measured by the number of narrative components
(4) There are qualitative differences in narrative coherence between grades as measured by differences in frequency among the individual narrative components
(5) Participants' stories become more cohesive from first to fourth grade as measured by the number of cohesive devices
(6) There are no qualitative differences in cohesion between grades as measured by the frequency order of the subcategories of cohesion (lexical ties > references > connectives > ellipses and substitutions).

No hypotheses will be put forward with respect to an increase or decrease of the individual categories of cohesion from first to fourth grade, i.e. whether qualitative differences in cohesion are to be expected as measured by each category's number of ties in first and fourth grade. Also, no hypothesis will be formulated with respect to

a relationship between coherence and cohesion due to the controversial findings of previous studies.

4 Research Design

4.1 Participants and data collection

4.1.1 The Kiel project

In Germany, the provision of foreign language teaching differs between the sixteen different states (*Bundesländer*), since responsibility for the education system is determined by Germany's federal structure.[1] In Schleswig-Holstein, Germany's northernmost *Bundesland*, a project called into life by Henning Wode from Kiel University seeks to enable participants to reach "meaningful communicative competence in at least two other languages in addition to his or her mother tongue" (Commission of the European Communities 2003: 4) as part of their formal schooling – something EU policy makers have called for on various occasions (e.g. also Council of Europe 1998). Since this requires a coordination of language teaching in primary and secondary education, the Kiel Immersion (IM) Project combines a bilingual preschool, an IM elementary school and bilingual wings at secondary schools. The present study deals only with data from the elementary school of the project and thus only this part of the project will be described in detail in the following. However, a short overview of the preschool program will also be given, since some of the participants in the present study attended the bilingual preschool beforehand.[2]

The bilingual preschool of the project (ages 3-6) was set up in Altenholz near Kiel in 1996 (cf. Wode 2009 & 2000, Burmeister & Pasternak 2004, Kersten et al. 2002, Rohde & Tiefenthal 2002). At least one group per year is attended to by a team

[1] Children generally enter primary schooling (elementary school) at age 6 to 7. Secondary school starts at age 10 to 11. Children usually attend preschool (age 2-3 to 6-7) before entering primary school; this is, however, not legally binding. For more details on the German education system as well as variation between the different *Bundesländer* cf. KMK (2009).

[2] For the secondary school part cf. e.g. Burmeister & Daniel 2002, Wode 1998, 1995 &1994. For a short overview of the Kiel project cf. also Möller 2009.

of a German and an English-speaking caregiver. By 2005, the preschool had three caregivers with English as their L1 and three of the five preschool groups were educated bilingually.[3] Bilingual education in the preschool follows the 'one person-one language principle', i.e. German- and English-speaking caregivers consistently use only their respective L1 when speaking to the children.

The neighboring elementary school in Altenholz started a partial English IM program in August 1999 (cf. Wode 2009, Burmeister & Pasternak 2004, Wode et al. 2003, Kersten et al. 2002). Regularly, one IM class per school year is offered, even though two first grade IM classes started in the 2005/2006 school year due to high demand.[4] The IM elementary school continues L2 English education for children from the bilingual preschool, but it is also open to children from monolingual preschools. All subjects except German language arts are taught in English, i.e., 70% of the curriculum, while the academic curriculum corresponds to that of the monolingual peer classes. In first grade, the children are allowed to speak German, although the use of English is highly encouraged; from second grade on the children are expected to use English only. Formal correction of errors and grammar explanations are kept at a minimum during the teaching of the academic subjects. Similarly, subject matter tests are taken in the L2 even though only the academic content is graded.

Due to the German teacher training system, all IM teachers in Altenholz have a teaching certificate for elementary and secondary school,[5] although they specialized in English for secondary school. Native speaker input comes from assistant teachers[6] and from a changing number of English L1 students whose parents work in Kiel for some time. Most of the teaching material has to be designed by the teachers, e.g., by translating German text books or creating new teaching materials with the help of pictures and corresponding English texts, since L1 English text books are linguistically too difficult at the beginning and do not correspond to the German academic curriculum.

All parts of the Kiel project have been evaluated by Henning Wode and his group at Kiel University. The preschool project was evaluated from 1997 to 2001 (e.g. Tiefenthal 2009, Wode 2009, 2004 & 2000, Rohde 2005, Rohde & Tiefenthal

[3] Personal communication Mrs. Devich-Henningsen (head of the preschool)
[4] The success of the IM project in Altenholz also inspired an increasing number of elementary schools in Schleswig Holstein and the neighbouring *Bundesland* Hamburg to set up immersion wings following the Altenholz model (cf. http://www.fmks-online.de, for example).
[5] I.e. *Grund- und Hauptschullehrer.*
[6] One assistant teacher per year is shared by all grades. (S)he teaches 14 hours per week.

2002). Research results show that even at the end of preschool, the children hardly use the L2, since there is no real necessity: All caregivers, even the ones with L1 English, understand German. Nevertheless, the L2 input in preschool is important, since it is responsible for the development of strong receptive abilities.

Research results at the elementary school show the following general developmental pattern for the L2 English (e.g. Kersten 2009, Wode 2009, Wode et al. 2003, Burmeister & Pasternak 2004, Kersten et al. 2002): During the first year of elementary school there is an impressive rise in the children's productive abilities;[7] verb morphology is dominated by *–ing* forms. At the end of second grade predominantly simple past forms are used and overgeneralizations like *felled* show that the general rule for the formation of the simple past has been learned. At the end of third grade many verb forms are produced target-like, even many irregular past forms, e.g., *fell*, *took*. At the end of fourth grade few non target-like verb forms remain. The children's stories have become longer and their syntax quite complex; a broad range of conjunctions and function words is used. The children are able to communicate freely and efficiently in English. L1 tests at the elementary school show that at the same time the L1 suffers in no way (e.g. Wode 2009, Zaunbauer et al. 2005, von Berg 2005, Bachem 2004, Wode et al. 2003).

While the elementary school students' linguistic proficiency and development in areas such as syntax, verbal morphology or phonology has been covered by quite a number of studies, children's discourse abilities have not yet been investigated on a larger scale. That is, only two studies have been conducted on discourse production measures and both of them on cohesion (Möller 2003, Maschewski 2002). No attempt has so far been made to study coherence and/or the relationship between coherence and cohesion. The present study seeks to fill this gap.

4.1.2 Participants

The data used in the present study is a subpart of a larger amount of data collected in the Kiel Immersion Project since 1999 (cf. Wode 2009 and the previous section). The learner variables collected as part of the project were sex, age in years, and L2

[7] Especially the receptive abilities acquired in preschool seem to form the basis for the rapid increase in productive abilities in grade 1 when classroom activities strongly encourage the use of the L2.

preschool experience. Other measures, such as parental demographics or socio-economic data, were not collected.[8]

Three experience groups are distinguished with respect to L2 preschool experience: Children with previous L2 English experience from a bilingual preschool ('bili' group),[9] children who attended a regular monolingual German preschool ('mono' group), and children who went to a monolingual German group in a preschool offering also bilingual groups ('MB' group). Since the latter preschool had an "open concept," i.e. children were relatively free to move around among groups, previous L2 experience can neither be confirmed nor ruled out for the MB group. At the same time the MB group is very small (N=3 in first grade and N=4 in fourth grade) and its usefulness for separate analyses is thus strongly limited. These limitations were kept in mind for the analysis: The MB group was included in the statistical analysis on all overall measures, but not as a separate experience group. Consequently, it was excluded from any statistical analyses on the influence of prior L2 experience.

The present study makes use of data collected at the end of the first and at the end of the fourth school year, i.e. after approximately 9,5, respectively 41 months of exposure. It combines longitudinal data from the first cohort of participants[10] in the IM program with cross-sectional data from the fifth and eighth cohort.

An overview of the different cohorts is given in table Tab. 4.1. All participants (N=66) have German as their L1 or stronger language and, according to the teachers, all of them were normally developing. Cohorts 1 and 5 were taught by the same, cohort 8 by a different teacher.

[8] However, see Zaunbauer and Möller (2007) for data from other cohorts of the same program. According to their teachers, the socioeconomic status of all cohorts used in the present study is roughly comparable (aside from some minor interindividual differences).

[9] The children of the bili group attended the Altenholz bilingual preschool described in the previous section. It should be noted, however, that the exact amount of experience as well as information about the exact time of exposure spent in the bilingual preschool group were not tested or collected. It is one of the limitations of the present study that children with bilingual preschool experience are treated as one homogeneous group simply based on the fact that they had some exposure to English in preschool, while there might well be interindividual differences attributable to distinct amounts of time of exposure.

[10] Cohort 1 children who had been to the bilingual preschool were already subjected to tests in an earlier part of the Kiel IM project (e.g. Rohde 2005, Rohde & Tiefenthal 2002, Wode 2000).

Cohort (Grade)	Total N	Female	Male	B	M	MB	Mean Age (SD)	Start (school year)	Tested
1 (1)	16	11	5	8	5	3	6;8 (0,40)	1999	May 2000
1 (4)	16	11	5	8	5	3	9;8 (0,40)	1999	May 2003
8 (1)	18	11	7	11	7	-	6;8 (0,43)	2006	June 2007
5 (4)	16	11	5	8	7	1	9;8 (0,45)	2003	May 2007
Total/ Mean	**66**	**44**	**22**	**35**	**24**	**7**	**8;2 (1,6)**	**-**	**-**

Tab. 4.1 Descriptive statistics of the study's participants. "B" denotes children with bilingual preschool experience ("bili" group), "M" children without L2 preschool experience ("mono" group) and "MB" children who went to a monolingual German group in a preschool offering also bilingual groups. "SD" denotes the standard deviation.

In terms of participants' exposure time to the L2 in first and fourth grade, respectively, cohort 1 had an exposure time of approximately 9 months and cohort 8 an exposure time of approximately 10 months at the end of first grade.[11] That is, the mean time of exposure for the first grade children is 9,5 months. The mean time of exposure in fourth grade was 41,5 (cohort 1) and 40,5 months (cohort 5).[12] This makes for a mean exposure time of 41 months in fourth grade.

As table Tab. 4.1 shows, the cohorts of the longitudinal and the cross-sectional data-set are comparable in terms of all measures, i.e. mean age range, standard deviation and subgroups. This comparability justified combining the different data-sets for the present study. To ensure further comparability, possible differences between the longitudinal and cross-sectional data sets were explored for all overall measures of coherence and cohesion[13] as well as for randomly selected individual variables. No significant differences were found.

In the following, the descriptive statistics will be given for the individual cohorts which participated in this study. Since several children originally tested had to be excluded from the analysis, this will also be discussed briefly. Table 4.2 gives the participant information collected for children from cohort 1.

[11] Cohort 1 entered school at the end of August 1999 and was tested in mid-May 2000, Cohort 8 entered school in mid-August 2006 and was tested in mid-June 2007.
[12] Cohort 5 entered school in mid-August 2003. For both cohorts three summer vacations, which amount to 6 weeks per year in the German school system, were deducted. Other vacations and holidays are more variable and were thus disregarded.
[13] I.e. total number of narrative components, narrative index score and total cohesive density (total number of ties per clause)

Participant No. (C1-G1- ... or C1-G4- ...)[14]	Sex	English experience	Age in grade 1 (May 2000)	Age in grade 4 (May 2003)
1	Male	B	7	10
2	Male	B	7	10
3	Female	M	7	10
4	Female	M	7	10
5[15]	Female	MB	6	9
6	Female	B	7	10
7	Female	M	7	10
8	Female	M	6	9
10[16]	Female	M	7	10
11	Male	MB	7	10
13	Male	B	7	10
14	Female	B	7	10
15	Female	B	7	10
16	Female	B	7	10
17	Female	B	7	10
18	Male	MB	6	9

Tab. 4.2 Participants of cohort 1. "B" denotes children with bilingual preschool experience ("bili" group), "M" children without L2 preschool experience ("mono" group) and "MB" children who went to a monolingual German group in a preschool offering also bilingual groups. "SD" denotes the standard deviation.

Two children of cohort 1 were not included in the analysis, child 9 since it left the class after grade 2 and child 12 due to interviewer instructions in grade 1, which could have led to a biased understanding of the task, namely as a picture description task.

Table 4.3 gives the participant information collected for children from cohort 5.

Participant No. (C5-G4- ...)	Sex	English experience	Age (May 2007)
3	Male	B	10
6	Female	B	10
7	Female	B	9
8	Female	B	10
9	Female	M	10
10	Male	M	10
11	Female	M	9
12	Female	M	10
14	Female	B	9
15	Female	B	9
16	Male	MB	10
17	Female	M	10
18	Female	B	10
19	Male	M	10
20	Female	B	10
21	Male	M	10

Tab. 4.3 Participants of cohort 5 (grade 4). "B" denotes children with bilingual preschool experience ("bili" group), "M" children without L2 preschool experience

[14] In the text, references to individual participants are coded as follows: Cohort-grade-participant number. 'C1-G1-1', for example, refers to child 1 from cohort 1 in grade 1, 'C1-G4-1' to the same child in grade 4 (longitudinal data set).

[15] For child C1-5 some input of (a) further language(s) including English cannot be excluded, since her father is from an African country and was reported to also use English at home, albeit irregularly (cf. also Kersten 2009).

[16] Child C1-10 had a bilingual Polish-German background (cf. Kersten 2009).

("mono" group) and "MB" children who went to a monolingual German group in a preschool offering also bilingual groups. "SD" denotes the standard deviation.

Five children of cohort 5 had to be excluded from the analysis: Child 13 because he had lived in the United States for 4 years, children 23 and 24 because they had joined the class only in 3rd grade, and child 5 because he did not do the B-test. Child 22 had to be excluded due to biased interviewer instructions (similar to child 12 in cohort 1). Additionally, three children who had been tested in grade 1 had left the class: two moved away and one skipped second grade.

Table 4.4 gives the participant information collected for children from cohort 8.

Participant No. (C8-G1- ...)	Sex	English experience	Age (June 2007)
1	Male	B	7
2	Female	M	7
3[17]	Male	B	7
4	Female	B	6
5	Male	B	7
6	Female	M	7
7	Male	M	7
8	Female	M	6
9	Female	B	7
10	Female	M	7
11	Male	B	6
13	Female	B	7
14	Female	B	7
15	Male	B	7
16	Female	M	6
18	Female	M	7
20	Male	B	7
21	Female	B	7

Tab. 4.4 Participants of cohort 8 (grade 1). "B" denotes children with bilingual preschool experience ("bili" group), "M" children without L2 preschool experience ("mono" group) and "MB" children who went to a monolingual German group in a preschool offering also bilingual groups. "SD" denotes the standard deviation.

Five children of cohort 8 also had to be excluded. Child 12 refused to do the B-test after a very straining A-Test. Child 17 had to be excluded due to suggestive interviewer contributions. Child 19 and 23 had both lived in the US for a prolonged period of time (4 ½ and 1 year, respectively) and were thus not included in the analysis.[18] One child (22) refused to participate in the interviews altogether.

In addition to the background variables grade, age, sex, and experience group, a short interview was conducted with cohorts 5 and 8 before the actual storytelling

[17] For C8-3 some input of (a) further language(s) including English cannot be excluded, since his father is from an African country (cf. C1-5).
[18] Additionally, child 23 had joined the class just two days prior to the interviews.

task about their experience with the story genre and storytelling.[19] All of the children reported at least some experience with stories or story-like genres ranging from respective TV shows to parents' bedtime stories or the children's own reading. Amount and especially type of experience varied especially as a function of age.

In Schleswig-Holstein different text genres and especially stories are also part of the curriculum for L1 German classes throughout primary school (Ministerium für Bildung 1997: 51ff.). This ranges from comprehension activities such as reading a book in class and talking about its content to students' own text production, e.g. in the form of oral (and later on also written) personal or fictional narratives, as well as activities centring on constructing texts from sentences.[20] Consequently, it is to be expected that apart from their varying home experience with stories all participants of the present study have had a significant school exposure to narrative structures and cohesive devices; this relates especially to the fourth graders.

4.1.3 Data collection: Materials and procedure

From 2000 to 2007 picture-elicited oral storytelling tasks for L1 German and L2 English were administered once a year at the elementary school of the Kiel IM Project (e.g. Wode 2009). Participants' L1 and L2 were tested on separate days with the L2 tests being conducted before the L1 tests. For all tests children were collected in their respective classrooms during regular teaching hours and tested individually in a separate room at the school. Since the present study deals exclusively with participants' L2 data, only this part of the testing will be described in the following.

The picture book *Frog, where are you?* (Mayer 1969; henceforth 'frog story'), which has been used successfully in quite a number of studies (e.g. Akinçi et al. 2001, Severing & Verhoeven 2001, Berman & Slobin 1994, Bamberg & Marchman 1991), was used as elicitation material for the L2 data. In black-and-white pictures the book tells the story of a little boy and his dog, who embark on a search for a lost frog (cf. ch. 4.2.1 for a detailed description of the story's structure).[21]

L2 data collection involved two different parts, an 'A-test' and a 'B-test', which were conducted right after one another on the same day. The test design serves to

[19] These interviews have not been evaluated systematically so far, i.e. no systematic quantitative analysis has been carried out, so that the results given here need to be considered somewhat impressionistic.
[20] According to the four cohorts' teachers, however, systematic activities centering on texts and especially text production are to be expected mainly from second grade on.
[21] For additional information on the frog story cf. especially Berman & Slobin 1994.

optimize conditions for participants' story production. That is, the A-test serves as a trial run to allow participants to optimize their performance in the B-test. Out of this rationale the order of A- and B-test was not counterbalanced[22] and only the B-test stories were analyzed in the present study.

Both A- and B-test have in common that participants were first of all encouraged and given sufficient time to look through the elicitation material in order to allow the formation of a mental representation of the events depicted. Additionally, children were allowed to look at the pictures while telling the story in order to avoid problems attributable to working memory constraints rather than cognitive or linguistic processing. Differences in the set-up and function of the two tests will be described in the following.

In the A-test the interviewer was known to the participants as a speaker of German and English. Participants were instructed to look through the story carefully and then tell the story to the interviewer in English. At the same time the children were told explicitly that the interviewer could help with vocabulary questions if requested. Interviewer instructions were to help only when asked and to otherwise give exclusively non-specific prompts encouraging storytelling. In the A-test interviewer and participant looked at the pictures together during storytelling; the children turned the pages at their own pace. The A-test thus serves two main purposes. First of all, it familiarizes participants with test situation and elicitation material. Secondly, the children are supported linguistically while they make a first attempt at encoding their mental representation of the story content. Since the interviewer can help to eliminate temporary linguistic difficulties, participants' linguistic processing load is lessened.

In the B-test participants were asked to tell the same story to a second interviewer, who they believed to speak and understand English only. The children were given time to look through the picturebook once more before the B-test. During the test they were again allowed to look at the pictures while telling the story and to turn the pages at their own pace. However, participants were instructed not to let the second interviewer see the pictures in the B-test. The interviewer in turn was instructed not to let the child show her or him the pictures and to give only non-specific prompts. In the B-test participants can focus on optimizing their linguistic expression of the content, since the cognitive load of conceptualizing the story has

[22] Similarly, L1 and L2 tests were not counterbalanced because this would have given children telling the story first in the L1 and then in the L2 two trial runs.

greatly been reduced through the A-test. That is, children have had sufficient opportunity to form a coherent mental representation of the story. Additionally, they have also had an opportunity to practice the linguistic encoding of this representation. However, in the B-test the children also need to cognitively and linguistically process an information gap between them and the interviewer, i.e. they have to modify their linguistic encoding according to the needs of an uninformed listener.

All interviews were recorded and then transcribed verbatim using the following transcription conventions:

1, 2 etc.	Participant numbers
#	short pause (500 msec- 5.0 sec)
##	long pause (5.5-10 sec)
###	very long pause (> 10 sec)
< >	transcriber comments
x	incomprehensible (roughly approximating the number of syllables)
/	speaker interrupts or corrects his or her utterance
italics	German words
,	slightly rising or falling intonation (as usually before a subordinate clause)
.	strongly falling intonation (as usually in signalling the end of an utterance)
?	strongly rising intonation (as in signalling a question)
!	intonation corresponding to an exclamation (e.g. *frog!*)
:	sound prolonged (e.g. *and the:n, psh:t!*)
ehm	any type of hesitator (German *äh*, *hmm* etc.)
hehe	laughter
|	Overlap (more than one person speaking at the same time)
B	B-test
I	Interviewer known to the children as speaking both German and English (A-test interviewer; may still be in the room and interact in the B-test)
IE	Interviewer known to the children as a native speaker of English, who does not understand or speak German (B-test interviewer)

4.2 Method of analysis

4.2.1 Narrative coherence: The structure of "Frog, where are you?"

As Fig. 4.1 shows, the task material "Frog, where are you?" (Mayer 1969), i.e. the 'frog story', has an excellent fit with the prototypical episodic structure described by story grammars (cf. ch. 2.3.2). In the following, I will briefly outline the realization of each of the grammar's terminal nodes in the frog story and how the analysis of these components was conducted.

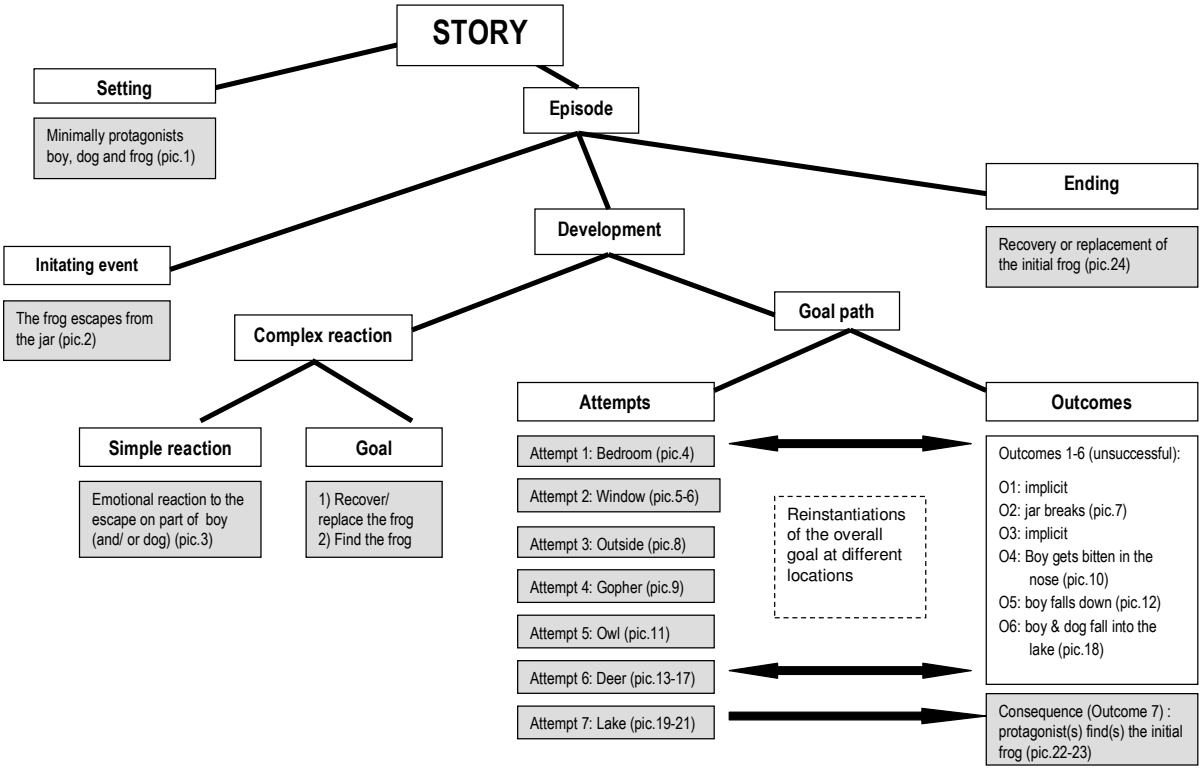

Fig. 4.1 The structure of the frog story in narrative components as identified by story grammars. Note: Terminal nodes taken into account for the analysis are shaded in grey.

Picture 1 of the frog story provides the SETTING: At night, a boy and a dog are sitting in a bedroom; they are looking at a frog in a glass jar. As explained in chapter 2.3.2, the minimal requirement for a SETTING in any story is the mention of one animate character, i.e. the protagonist. In the frog story this corresponds to three protagonists: Boy, dog, and frog.[23] Thus, participants were credited with the SETTING component, if they made explicit reference to all three protagonists at the beginning of the story.[24]

The INITIATING EVENT follows from the second and/ or third picture, namely the frog's escape (e.g. Reilly et al. 1998). In the task material, picture 2 shows the frog climbing out of his jar, while the boy and the dog are sleeping. Even though the

[23] The frog story contains several characters with their individual perspectives and thus several story strands (cf. Trabasso & Rodkin 1994: 91). The two main story strands, which overlap often but not always, relate to the boy and the dog (cf. Bamberg & Marchman 1994: 558f., Bamberg 1987: 23). The narrative components investigated in this study are based on the boy's experiences, goals, adventures etc., since he is the main protagonist (cf. Trabasso & Rodkin 1994: 91).

[24] Participants could also have opted to tell the story as having only two protagonists, e.g. boy and frog or dog and frog, in which case the mention of only two protagonists in the SETTING would have been systematic and thus acceptable. However, all participants mentioned the three protagonists at some point in their storytelling – even if this was not the case in the SETTING.

window is depicted as partly open, the actual escape of the frog has to be inferred: In picture 3 daylight is coming in through the window and the boy and the dog are lying on the bed looking at the empty jar; the boy's face indicates an emotional reaction to the empty jar. In order to receive credit for the INITIATING EVENT, participants needed to make reference to either the frog escaping, the result of this action, i.e. the jar being empty or the frog being away (picture 3), or the boy's discovery of the escape (cf. Berman & Slobin 1994: 52f., Kupersmitt & Berman 2001).

Participants were credited with reference to the boy (and/or dog) noticing the escape if they used a verb of perception or cognition - the clearest case being a verb of perception, e.g. *see* in example 4.1. The verb *look,* in contrast, can refer to both a physical and a mental activity, even if its core meaning relates to the domain of physical activity (Biber et al. 1999: 361). Therefore, a form of *look* was only accepted as encoding the discovery, if a description of the result of the escape followed, e.g. in example 4.2. In cases where *look* was preceded by a description of the result as in example 4.3, though, it more likely indicates the beginning of the search and was thus not credited as discovery but as an ATTEMPT (see below). Another option considered a realization of the discovery was the use of direct speech such as in example 4.4, which indicates the result of a thought process.

4.1 5[at the next morning, the little boy saw] 6[the glass was empty.] (C1-G4-1)

4.2 4[and the frog # is # go out.]… 5[And # on/ # on the da/ day the # boy looks/ # ehm looks,] 6[and the frog are # not there.] (C1-G1-5)

4.3 3[the frog run away.] # 4[tomorrow, the frog are not here.] # 5[the boy # look to the frog.] (C8-G1-14)

4.4 3[the <de> frog # hopp out the </de/> ### glass.] # 4[the boy's shout] 5[where is my frog?] (C8-G1-21)

The narrative component SIMPLE REACTION, i.e. an internal response, follows from picture 3, in which the boy's face clearly indicates an emotional reaction, even if the exact emotion, e.g. surprise or sadness, is open to interpretation. Consequently, the simple reaction was considered realized if explicit reference was made to the boy's (and/ or dog's) emotional or cognitive reaction to the frog's escape, i.e. either the description of an emotional state as in example 4.5 or the reference to a thought process as in 4.6. A third possibility accepted were interjections such as *oh no* in example 4.7, which have the sole purpose of expressing an emotion. Exclamations such as in 4.8, on the other hand, are the result of a cognitive process but do not

include any direct linguistic expression of a thought process or emotion. Thus, they were considered part of the INITIATING EVENT (discovery).

4.5 11[the boy was very sad.] (C1-G4-5)

4.6 6[in the morning, 7[when he woke up,] he looked to his glass,] 8[and thought,] 9[where is my frog?] 10[is he gone?] (C5-G4-19)

4.7 3[then the little children # stand up] 4[and # sayd … # oh no!] # 5[where's I/ <?> the frog!] ## 6[fro-og, where are you?] (C8-G1-20)

4.8 3[the <de> frog # *hopp* out the </de/> ### glass.] # 4[the boy's shout] 5[where is my frog?] (C8-G1-21)

Two GOALs can be assumed to motivate the goal path and thus a globally organized narration of the frog story: On the one hand the goal "recover or replace the frog" and on the other hand the goal "find the frog" (cf. Trabasso & Rodkin 1994: 94)[25]. Either GOAL by itself is enough to act as a global organizing theme for the frog story (cf. Berman & Slobin 1994: 46ff., Bamberg & Marchman 1991: 281), even if the pictures clearly suggest that merely finding the frog is not the actual end of the story. If both goals are realized in a narration, the higher-order goal "recover or replace the frog" (GOAL 1) motivates a secondary goal "find the frog" (GOAL 2), which in turn motivates attempts in particular locations (Trabasso & Rodkin 1994: 94. [26]

The lower-order goal "find the frog" (GOAL 2) was considered realized explicitly, if participants used a verb or phrase expressing an intention, desire or strategy to find the frog as part of the ATTEMPT section[27], e.g. an adverbial *to*-phrase expressing purpose[28]. Thus, examples 4.9 and 4.10 were considered a realization of GOAL 2, while 4.11 was coded as an ATTEMPT:

4.9 15[the dog ehm put his muzzle into the glass jar,] 16[and wanted to look] 17[if there was any frog in ere/ in there.] (C1-G4-16)

4.10 38[now, they go to the wood, and ehm look/ ehm to look for Bingo.] (C1-G4-14)

4.11 11[He looked in ea/ each sh/ shoe, and everywhere.] 12[the dog put his hehehe head in the glass,] # 13[but he can't find him.] (C5-G4-19)

[25] Trabasso & Rodkin (1994) postulate only the higher-order goal "get the frog back". For the present analysis, this higher-order goal was seen as too restrictive, since different interpretations of the ending are possible: Recovery of the original frog or replacement by a new one (e.g. Berman & Slobin 1994). Thus, even though the differentiation between higher-order goal and subgoal is taken over from Trabasso & Rodkin, their higher-order goal is extended to "recover or replace the frog."

[26] More specifically, the secondary goal "find the frog" motivates local goal-plans, i.e. the goal "find the frog" is reinstated in particular locations, thus motivating local attempts (Trabasso & Rodkin 1994: 94).

[27] I.e. after the initiating event (cf. Fig. 4.1).

[28] Either overtly marked by *in order* or not (cf. Biber et al. 1999: 827):

The same criteria were applied for an explicit realization of the higher-order goal "recover or replace the frog" (GOAL 1), i.e. participants were given credit if they used a verb or phrase expressing an intention, desire or strategy to recover or replace the frog.

Goals can also be realized implicitly, though, in that they need to be inferred from a combination of other narrative components (cf. ch. 2.3.2). GOAL 1 was therefore considered realized implicitly if participants realized the ENDING, i.e. encoded a recovery or replacement of the frog at the end of the story, in addition to INITIATING EVENT and at least one ATTEMPT, as in example 4.12. GOAL 2 was considered realized implicitly, if participants encoded the INITIATING EVENT and at least two ATTEMPTs, thereby indicating a 'sustained search' (e.g. Akinçi et al. 2001, Trabasso & Rodkin 1994, Berman 1988).

4.12 3[at night 4[when the boy s/ slept] the frog crept out of his glass] 5[and 6[when the boy wakes up] the/ # the frog was away.] … 7[the boy looked in his boots,] 8[the # dog in the glass] 9[and then the dog had his head in the glass] 10[and he couldn't ge/ # getch out of the glass.] … 11[And then the dog fell out of the window] # 12[a:nd: the boy get out] 13[and # cried] 14[frog, where are you,] … 27[And then the boy took one of the/ of the children] 28[and # got away, at home.] (C1-G4-2)

The distinction between explicit and implicit realizations of GOAL statements is actually scalar: In addition to linguistic expressions directly encoding the main character's desire, intention etc., various other possibilities were exploited by the children to indicate the underlying goal(s), for example the frequently used exclamation or question *frog where are you* (subgoal), exclamations such as *please come back* (higher-order goal) or outcome statements such as "but he [the boy] can't find him [the frog]" (C5-G4-19). Since these expressions involve at least some inference on the part of the listener, though, they cannot be considered explicit linguistic realizations of the protagonist's goal(s). Thus, they were also coded as implicit realizations, but only if they had not already been given credit as ATTEMPT or OUTCOME (see below).

The GOAL is followed by a plan application sequence consisting of several ATTEMPTs to find the frog in different locations and their OUTCOMEs, i.e. a recursive goal-path. Only the very last attempt is successful; all other attempts inside and outside the house are met with failure. Following Bamberg & Marchman (1990, 1994), seven attempts were identified in the picture book (cf. Fig. 4.1).

The initial attempt (ATTEMPT 1), which instantiates the search (cf. Bamberg & Marchman 1994: 563), is shown in picture 4 ("bedroom attempt"): The boy is looking into a boot, while the dog has its head in the glass jar the frog had been kept in.[29] Participants were credited with the first attempt, if – in relation to picture 4[30] – they either made explicit reference to the search on the part of the boy or the dog (example 4.13), mentioned activities not depicted, e.g. *look* + *T-shirt* (C5-G4-16), or used generalizations of search activities depicted or not depicted, e.g. *look* + *all over* as in example 4.14 or *look in* + plural noun as in 4.15 (cf. Bamberg & Marchman 1990: 80ff.).

4.13 11[the dog search/ chu/ searches the frog in the glass,] # 12[and the boy searches him in/ in the:/ # a boot.] (C1-G4-18)

4.14 12[then # they looked all! over, in the boo:ts, in the gla:ss,] (C1-G4-13)

4.15 11[He look in he # s <his> boots,] # 12[and the frog are not there.] (C1-G1-6)

4.16 9[The boy looking in/ in the <but cf. A!> # shoe,] 10[and the dog in a glass.] (C1-G1-14)

Look + preposition (*in/ into/ to/ out* etc.) was only credited as realization of ATTEMPT 1, if additional mention was made of the frog or generalizations of the search such as described above, since otherwise this corresponds to a mere picture description as in example 4.16. The same is true for expressions involving *shout, scream etc.* only (e.g. C1-G1-11, clause 13f.).

Failure of the initial attempt (OUTCOME 1) needs to be inferred from the continuation of the search in picture 5, where the boy shouts out of the window ("window attempt", pic. 7). This second attempt also fails; it ends with the boy and the dog outside below the window with the jar broken into pieces. The search then continues outside ("outside attempt", pic. 8) and in the woods: The boy shouts into a hole in the ground and gets his nose bitten by a gopher ("gopher attempt", pic. 9-10), he shouts into a hole in a tree and falls down from the tree as an owl comes out of the hole ("owl attempt", pic. 11 & 12). After that the boy climbs on a rock and is carried away by a deer, which drops him into a lake ("deer attempt", pic. 13-17). Finally, the protagonist(s) make(s) an attempt to find the frog behind a log in the lake

[29] Picture 4 is the most likely choice to signal the instantiation of the search. Nevertheless, there are other acceptable options: Bamberg & Marchman (1990: 90), for example, found that several of their participants encoded picture 4 as a dressing scene and picture 5 as the instantiation of the search.

[30] If participants' utterances were not clearly attributable to picture 4 or 5 (e.g. C1-G1-02), they were coded as ATTEMPT 1, i.e. as referring to picture 4.

("lake attempt", pic. 19-21). This last attempt is finally met with success and thus completes the overall search theme (cf. Berman & Slobin 1994: 46, Bamberg & Marchman 1990: 74): The protagonist(s) find(s) two big frogs and several smaller frogs, i.e. a "frog family" (pic. 22 & 23).

All attempts succeeding the instantiation of the search, i.e. ATTEMPTs 2 to 7, were coded following the same guidelines as ATTEMPT 1. OUTCOMEs 1 to 6 were not considered since they follow automatically from their corresponding attempts – even if no explicit mention of failure is made – e.g. due to a subsequent attempt (example 4.17) or because the child simply continues with the storytelling, and they serve no additional function in the story.

4.17 13[and they open the window] 14[and shout/ shouted,] 15[frog, where are you?] … 22[and they got to the forest] 23[where they shouted] # 24[frog! <fro-og> where are you?] (C1-G4-4)

The CONSEQUENCE (also OUTCOME 7) was coded, however, due to its function of concluding the search theme. Participants were credited with the CONSEQUENCE (OUTCOME 7) if they identified the frog as being the one from the beginning of the story – either in relating the content of pictures 22 and 23 (explicit realization) or as part of the ending (implicit realization).

The ENDING of the story is shown in picture 24: The boy and the dog are walking away from the frog family; the boy has a frog in his hand and is waving good-bye. Participants were credited with the ENDING component, if they made explicit reference to either the recovery or replacement of the original frog, i.e. the boy either takes his initial frog with him or a new frog to replace the one he lost. Consequently, participants were credited with realizing the recovery option, if they explicitly identified the frog as being the original one and made reference to the boy taking the frog away with him. The replacement option, on the other hand, was considered realized, if reference was made to the boy taking a frog away with him, which had either simply not been identified as the original frog (e.g. C1-G1-14) or explicitly introduced as a different one (e.g. C1-G4-2).

Two general problems occurred in the analysis of narrative coherence: First of all, as in any analysis involving spoken and especially spoken learner language, some degree of interpretation was involved in judging whether something was given credit or not. As a general rule, interpretation was kept to a minimum and critical

cases were discussed with other linguists.[31] A second problem in the analysis was the occurrence of German expressions in the L2 English data (cf. also ch. 4.2.2.1). Although narrative coherence, i.e. the realization of narrative components, is considered a cognitive phenomenon, it is only accessible via its linguistic expression side.[32] Thus, the central question is whether participants are able to express the narrative components in their L2. As a general rule, participants were not credited with a narrative component, if the crucial element of the clause was in German, as for the ENDING in example 4.18 and the INITIATING EVENT in example 4.19:

4.18 21[and then have the boy a little frog # *wieder*.] (C8-G1-2)

4.19 5[the glass # is *leer*,] (C8-G1-8)

4.20 3[and the frog # *riß* <?> *aus*/ out the glass.] (C8-G1-2)

They were given credit, on the other hand, if the narrative component could still be considered accomplished when leaving out the German expression, as for the INITIATING EVENT in example 4.20.

Sometimes it was even necessary to recur to intonation in order to decide whether to include or exclude a clause:

4.21 10[the boy are *ruft*, <engl. Aussprache> fro:g! fro:g!] (C1-G1-4)

4.22 10[and the boy *ruf*ing # the frog.] (C1-G1-7).

In the examples, C1-G1-4 (example 4.21) uses direct speech and thus conveys the sense of calling, even if the German verb is discarded. In the case of C1-G1-7's utterance (example 4.22) only a pointless combination of two noun phrases would remain.

4.2.2 Cohesion

The number of cohesive devices is to a large degree influenced by text length, since the number of ties increases with each additional clause. To exclude the influence of this factor the transcripts were first coded into clauses and then the proportion of cohesive devices per clause was used as a measure for cohesion. The rationale for the choice of the clause as the basic unit of analysis as well as the coding method

[31] Due to temporal and financial constraints no inter- or intrarater agreement was conducted; this is one of the limitations of the present study.

[32] A separation into linguistic, cognitive, and communicative abilities as well as their development is actually rather artificial and betrays the strong interconnectedness of these abilities, especially in relation to narration (e.g. Berman 2001b: 25, Bamberg 1987: 13; cf. also ch. 2).

are described in the next chapter. The method used for the analysis of cohesive devices is described in the ensuing chapters.

4.2.2.1 Coding into clauses

The transcripts were coded into clauses following the semantico-syntactic definition used by Berman and Slobin (1994): Any unit containing "a predicate that expresses a **single** situation (activity, event, or state), including finite and nonfinite verbs as well as predicate adjectives" (Berman et al. 1986 in Berman & Slobin 1994: 657).

The clause was chosen as the basic unit of analysis because it is a comparatively well-defined grammatical unit.[33] That is, clauses allow a rather precise allocation of cohesive devices and their respective anaphoric elements. At the same time a coding into clauses allows to measure participants' text length and thus to calculate the texts' cohesive density. An additional advantage of clauses is that they are semantically close to propositions, whose usefulness for studies of text production and processing has amply been demonstrated (e.g. Kintsch 1998 & 1974, Reilly et al. 1998, Mandel & Johnson 1984, Thorndyke 1977). While propositions are located on a conceptual, cognitive level, though, clauses are a linguistic unit realized in speech or writing; it can be said that clauses express propositions (cf. Cruse 2004: 22, Huddleston & Pullum 2002: 34, Biber et al. 1999: 122, Kintsch 1998: 54ff., Reilly et al. 1998, Mandel & Johnson 1984, Thorndyke 1977, Kintsch 1974: 5). Due to being centered around the verb as expressing one state, event or activity clauses thus also allow a rather precise allocation of the conceptual units of discourse realizing individual narrative components.

Following Berman & Slobin's coding (1994: 660ff.) for comparability, lexical verbs similar in function to modals and aspectual verbs, e.g. *want* or *start*, were coded together with the verbs they modified if they occurred in utterances with the same subject as in example 4.23. If, on the other hand, they introduced utterances with a different subject, they were coded as belonging to two different clauses (example 4.24). Similarly, narrator comments, i.e. metacomments on story content (example 4.25) or the communicative situation were coded together with the clause

[33] Especially compared to sentences (cf. Quirk et al. 1985: 47, Hunt 1965). For a discussion of the (dis-)advantages of several units of analysis particularly with regard to measuring text length cf. also Möller (2009a).

they commented on. If they formed a clause of their own they were coded accordingly but did not count towards total text length.[34]

4.23 19[because they wanted to find Spot.] (C1-G4-7)

4.24 50[and he wants] to/ # to/ 51[the deer to let the boy down.] (C5-G4-15)

4.25 66[and it look like # ehm the dog was ill,] (C5-G4-24)

Several coding problems arose due to the fact that this study looks at spoken language and, moreover, L2 learner data. These were unintelligible passages, German words in the L2 English transcripts, dysfluencies, verbal interaction with the interviewer, and verbless clauses. In the following, the coding procedure for each of these problem areas will be described.

Utterances with unintelligible passages were coded into clauses if only optional elements, i.e. modifiers (e.g. Quirk et al. 1985), were unintelligible and the overall meaning of the utterance could still easily be inferred (example 4.26). Otherwise these passages were not coded.

4.26 32[and find a Loch on X/ X.]< hit/ him ?> (C8-G1-1)

In the L2 transcripts, utterances with predominantly German words were not coded as clauses, since the question was whether the children were able to tell a coherent and cohesive story in English. Thus, utterances consisting to more than 50% of German words were discarded. First, the total number of words in the utterance as well as the total number of German words was counted. Articles, prepositions and conjunctions were not included and phrasal verbs treated as one unit. Words which were not completely target-like in their realization but sufficiently similar to their L2 targets, e.g. *been* or also *beens* ('bees', German *Bienen*), were treated as English words, even if they could also have been considered instances of code-switching to German. Words that were simply "anglicized" in pronunciation, on the other hand, were counted as German. Then, the number of German words in an utterance was divided by the total number of words in the utterance and the result was checked against the 50% threshold.

Dysfluencies such as (self-) repairs, repeats and hesitations (for the terminology cf. Biber et al. 1999) occurred very frequently in the data. They are a performance phenomenon that is typical for spoken language (cf. Biber et al. 1999: 1039) and even more so for learner language. Where simple repeats occurred within clauses,

[34] In this they were treated like crucial plot information, cf. below.

they were disregarded for the analysis. This is illustrated in example 4.27, where the child uses repeats and hesitators to extend processing time and indicate it has not yet finished the utterance (cf. Biber et al. 1999: 1054f.).[35]

4.27 5[that the frog is: # ehm # {the frog is:} # e:hm outgo # of the glass.] (C1-G1-18)

When retrace-and-repair sequences[36] occurred in the utterances, e.g. the self-repair in example 4.28, only the reformulation was coded. As a general rule, only clause elements that the speaker retraced explicitly were excluded from analysis. Thus, if the speaker repaired the connective in a retrace-and-repair sequence, the clause was coded from the new connective on, e.g. in example 4.29 where the simple sequential connective *then* is replaced by the more specific, punctiliar connective (cf. Halliday & Hasan 1976) *in this moment*.

4.28 # ehm, then he shaked the tree/ *also* 44[then the dog shaked the tree,] (C5-G4-21)

4.29 then/ ehm 42[in this moment # a/ all bees was/ were/ were following,] (C1-G4-13)

If, on the other hand, element(s) following the connective were repaired but there was no reformulation of the connective itself, the clause was still coded from the connective on. This was done since retrace-and-repair sequences involve "backtracking to the initial part of a clause" (Biber et al. 1999: 1063), e.g. a connective, as opposed to a complete abandonment of the original utterance.

Self-repairs sometimes led to utterances with apparently incoherent syntactic structure and some degree of interpretation was necessary to decide which parts of the utterance were repaired and what the final intended structure was. This process will be demonstrated for example 4.30: First, the speaker self-repairs the verb form *cames* and finishes the originally planned utterance with the particle *out*. The original structure of the utterance is left intact; the reformulation is accommodated in the existential clause structure.

4.30 52[then there cames an owl/ came an owl out/ # <atmet tief ein> flew out,] (C1-G4-17)

The next self-repair, though, allows for two interpretations:

[35] Utterance parts that did not enter the analysis are enclosed in {…}.

[36] The terms repair, recast, reformulation and retrace-and-repair-sequence will be used interchangeably for any sequence where part(s) of an utterance are reformulated by the speaker thereby "retracing" or cancelling the original part(s).

(1) The speaker not only reformulates the original verb into a more specific verb of motion but also retraces the existential clause structure. This means that a syntactic blend is created, i.e. the utterance starts with a different grammatical structure than it finishes with (cf. Biber et al. 1999: 1065). Only the connective would still be accommodated in the new structure. Thus, the clause would enter the analysis as [then an owl flew out].

(2) The participant repairs again only the verb, as a kind of afterthought; evidence for this are the short hesitation pause and the parallelism of *flew out* to the retraced construction *came […] out*. This would mean that the original structure remains intact and the clause enters the analysis as [then there flew an owl out].

Following the general rule established for retrace-and-repair sequences, namely to exclude from analysis only clause elements that the speaker retraced explicitly, interpretation (2) was chosen.

The interviews show a great awareness on the part of the participants, even in grade 1, as to the stereotypical roles in a storytelling context, i.e. an active storyteller and a comparatively passive listener. This set-up was encouraged through the interviewer instructions, of course, since interviewers were told to use non-specific prompts such as *mhm* or *okay* and in general to focus on the non-verbal aspects of their role as "fascinated listeners." Nevertheless, some verbal interaction between participants and interviewers was unavoidable – its complete absence would have been as unnatural as in any other communicative context. The question was then if and how to code utterances resulting from interaction with the interviewer. Interaction was sometimes initiated by the interviewer and sometimes by the participating children. Interviewer-initiated sequences resulted from prompts as well as from clarification requests. Participant-initiated interaction sequences, on the other hand, usually followed vocabulary questions.

First of all, interviewer-initiated sequences and their coding will be described. Prompts used by the interviewers can be categorized as being either non-specific or specific. Non-specific prompts are general interviewer contributions, which serve to encourage the child in telling the story and, if necessary, to encourage it to tell more; this includes non-verbal and verbal (e.g. *oh*, *mhm*) expressions of interest, surprise etc., as well as questions such as *What is happening?*. More specific prompts such

as *And how about the dog?* were used only if a child was very reluctant to speak at all. Additionally, the interviewers made use of non-specific and specific clarification requests. Non-specific clarification requests were, e.g. questions to repeat an utterance due to background noise or because the child had mumbled part of an utterance. Specific clarification requests, on the other hand, were direct reactions to something the child had said, e.g. a German word, a non-target-like structure or an obvious mismatch in content with the previous utterances (example 4.31).

Example 4.31

16 the frog <dog> ehm took his tongue a/ and # ehm

IE The frog or the dog?

16 *Ehm* 29[the dog took his tongue] 30[and the/ # wish/ wif/ the wif/ wish <from German wischen> ehm over his cheeks.] (C5-G4-16)

Children's reactions to non-specific and specific prompts were coded like normal utterances, except when they were mere repetitions of what the child had said before. For all other interaction sequences, coding was based on the children's verbal reaction – again excluding simple repetitions. When children's utterances did not function as an answer to an interviewer question, but rather seemed attributable to a lag in the child's processing of its originally planned utterance, they were coded like normal.

More often, though, interaction with the interviewer resulted in recasts. As a general rule, only the recast was coded, if it presented an "improved" version of the original utterance, e.g. by clarifying the actor or replacing a German with an English expression (cf. example 4.31 above and 4.32 below, where recasts follow a clarification request). In this case, the original utterance was not coded.

Example 4.32

9 … then # <they, cf. below> fall in the *Wasser*,

IE Wh/ what?

9 16[Then they </ði/> fall in the water,] (C8-G1-9)

Participant-initiated interaction sequences usually followed vocabulary questions and involved either a recast after successful negotiation for a vocabulary item[37] or a recast without any actual negotiation taking place. Passages with

[37] These negotiations took place in English. Interviewers were instructed not to answer German vocabulary question, since this would have made their role as monolingual speakers of English questionable.

reformulations resulting from successful negotiation were coded as follows: If the child managed to extract, e.g., an English word it was missing via negotiation and interaction with the interviewer, the interaction sequence between “missing piece” and unfinished clause was disregarded and the whole unit coded as one clause. This is shown in example 4.33, where {…} encloses sequences that did not enter into the analysis.

Example 4.33

6 34[and in one tree there was {a bee # heap. </hi/> haive? </eɪ/> a ha/ a have *oder? #* bee ehm/ # a bee # heap? # ehm why don’t/ it’s called, ehm a bee heath? </i/>

IE Oh, a beehive.}

6 A beehive.] (C1-G4-6)

Interaction sequences without an actual negotiation for vocabulary usually involved the child asking for a German word and the interviewer reacting with a general expression of incomprehension. The child then self-repaired and continued the original utterance replacing the German with an English expression (example Example 4.34). Again, the recast was coded and entered into the analysis, while the German expression was disregarded.

Example 4.34

10 27[and the dog # ehm # ehm # bellt?

IE I am sorry? I don’t understand.

10 X <bi/ ?> bark at the # beehive,] (C5-G4-10)

Verbless clauses were also coded. In this I follow again Berman & Slobin (1994: 661). Assuming an ellipsis was avoided, though, except for clear cases of gapping (cf. Huddleston & Pullum 2002: 1337f.) as in example 4.35 below. Thus, coordinated noun phrases, adverbs etc. were treated as one unit, since they syntactically function as one, even if semantically the respective predicate refers to both elements and two propositions could be argued for (example 4.36). This line of argumentation also holds true for existential clauses as in example 4.37, where a singular predicate is followed by two coordinated noun phrases as the notional subject. In written language this would be considered ungrammatical and an ellipsis would need to be assumed. In spoken language, however, the verb is often singular, even if the

following notional subject is plural (cf. Biber et al 2002: 236), so that these instances were also coded as one clause.[38]

4.35 7[the boy looked in his boots,] 8[the # dog in the glass] (C1-G4-2)

4.36 8[and on the ground lays a t-shirt and a sock.] (C1-G4-7)

4.37 1[Once there was a boy and a dog,] 2[they had a frog in a glass,] # 3[and they were/ # were/ were nice to the frog,] (C1-G4-4)

Verbless sentence elements introduced by *with* were coded as clauses if they had an S(V)A or S(V)C structure, i.e. a normal clause structure with verbal ellipsis, and if, additionally, they functioned as sentence adjuncts, i.e. grammatically separate from other clause elements (cf. Quirk et al. 1985: 478 & 511ff., Berman & Slobin 1994: 661). *With* functions as a subordinator in these cases and not as a preposition (Quirk et al. 1985: 705), compare e.g. examples 4.38 and 4.39.

4.38 81[the dog <boy> goes in the water again] 82[with the frog on his hand.] (C5-G4-9)

4.39 47[there they found two fro/ two big frogs with lit/ with many little small frogs.] (C1-G4-14)

Verbless clauses were also coded if verb or auxiliary were missing due to a deficient structure attributable to learner language, and if the meaning was easily recoverable by adding a form of copular *be* (example 4.40).

4.40 9[the dog falling down.] # 10[the glass broken.] (C1-G1-17)

Due to the nature of the storytelling task as an oral performance, the children sometimes "acted out" the verb instead of mentioning it explicitly. These cases were also coded as clauses (example 4.41). Finally, crucial plot information was coded even without the presence of a verb or an easily reconstructable ellipsis (example 4.42).

4.41 8[the boy, # fro:g! fro:g!] (C1-G1-4)

4.42 38[there was two frogs,] # 39[and # he/ the one was frogfrog.] # and he's/ <his> # 40[and he's babies.] (C8-G1-5)

4.2.2.2 Cohesion methodology

The present study's cohesion analysis was conducted based on the seminal work by Halliday and Hasan (1976, ibid. 1985, Hasan 1984). Thus, for each participant the

[38] At the same time, this does justice to Berman and Slobin's criterion of a unified predicate expressing a single situation (see above).

use of references, substitution, ellipsis, connectives and lexical cohesion was coded. The study takes into account intra- as well as interclausal ties in order to capture all possible links, even if interclausal (and even more so intersentential) ties are more salient in contributing to cohesion.[39]

All potentially cohesive items were extracted from the individual transcripts and coded according to (a) the subcategory of cohesion, e.g. reference or connective, and (b) the most immediate[40] presupposed item as well as the clause this presupposed item occurred in.

Additionally, each tie's phoricity was coded, i.e. whether references were endophoric or exophoric and anaphoric or cataphoric. Endo- and anaphoric reference was considered the "default" i.e. it was first of all assumed that a relationship exists with some preceding element in the text rather than looking for possible relationships with subsequent text parts. If both interpretations were possible, ties were coded as endo-/ anaphoric.

German words were excluded from the analysis even if they were "anglicised" in pronunciation or had an English homophone which was, however, contextually infelicitous, e.g. /fil/ in example 4.43, which could be interpreted as either German *fiel* or English *feel*.[41] Non-targetlike English words were coded if it was possible to infer the intended meaning due to closeness to the target word (example 4.44) or due to the linguistic context.

4.43 9[the dog *fiel* </fil/> *vor* <=from> the window.] (C1-G1-10)

4.44 18[and the beens*:* # fly after the dog,] (C1-G4-2)

After the classification into subtypes, which will be described subsequently, the following measures were calculated for each subcategory:

- Total number of occurrences
- Number of occurrences per clause (density).

Additionally, participants' overall cohesive density, i.e. the mean number of ties per clause, was calculated. Since only endophoric and anaphoric ties regularly contribute

[39] Intraclausal cohesive devices coded, such as for example possessives, were references and lexical ties; the connectives *and* or *but* as phrasal coordinators were not coded. The overall number of intraclausal device was very low.

[40] Cohesive items are very often part of cohesive chains involving repeated references to the same referent, several repetitions of the same lexical item etc. The degree of their cohesiveness is of course also determined through membership, i.e. embedding, in such a chain (cf. Halliday &Hasan 1976: 330). However, to preserve some degree of clarity in the analysis only the tie with the most immediate presupposed item is recorded.

[41] Indefinite or definite articles used together with German nouns, on the other hand, were included in the analysis of demonstrative references, e.g. the (*Eul*) (C8-G1-11).

to cohesion (Halliday & Hasan 1976), exophoric and cataphoric ties were excluded from this final step of the analysis.

4.2.2.3 References

References were classified according to the semantic relationship they encode. Thus, personal, demonstrative and comparative references were coded (Tab. 4.5).

Type	Subtype	Add. subtype	Example
References	personal	personal pronoun	*he, him, they, them, it…*
		possessive pronoun	*his..*
		relative pronoun	*who, which, that*
	demonstrative	adverb	*there, here*
		definite article	*the*
		demonstrative proper	*this, that*
		relative pronoun	*where (~in which)*
	comparative		*more, similar, other*

Tab. 4.5 Reference types included in the analysis

The category of personal references comprises personal (independent of case), possessive and relative pronouns. Left dislocations (also 'pronoun copies') such as in example 4.45 (cf. also Bamberg 1994: 226f., Biber et al. 2002: 956) were excluded from the overall count of (personal) references, since they are an oblique repetition.

4.45 5[The # boy, he's # very very # tired.] (C1-G1-3)

Since the present study's cohesion analysis is based on the clause, relative pronouns (e.g. *who, which*) were also considered. In contrast to interrogative pronouns (cf. Halliday & Hasan 1976: 309), relative pronouns contribute to cohesion in a clausal analysis, since they refer back to a referent in the text (cf. Quirk et al. 1985: 365), as in example 4.46.

4.46 1[once there was a bo:y,] 2[who had a f/ frog in a gla/ ehm ehm in a glass,] (C1-G4-6)

Similarly, possessive pronouns were coded as personal references even if their referent was in the same clause. In such cases the use of the possessive is strongly dependent on clausal structure and it could therefore be considered non-cohesive. Nevertheless, there is still a cohesion-building relationship of co-reference.

Classified as demonstrative references were the adverbs *here* and *there*, the determiners *this* and *that*[42], the relative pronoun *where* indicating location

[42] Demonstrative *that* is not to be confounded with *that* acting as a relative pronoun such as in: 3[and he had a dog] 4[that he called Lucky.] (C1-G4-07). In the example, *that* was coded analogously to *which* as a personal reference. Additionally, demonstrative and factual connective *that* were discriminated (cf. connectives).

analogously to spatial *there* (cf. Quirk et al. 1985: 442), and the definite article. The definite article *the* contributes to cohesion in that it indicates that "the noun it modifies has a specific referent, and that the information required for identifying this referent is available" (Halliday & Hasan 1976: 74). Thus, it signals that this information is recoverable either endophorically from the preceding text (anaphora) or from within the same nominal group (cataphora), e.g. *the hole of the owl*,[43] or on the other hand extralinguistically, e.g. in generalized exophoric use as in *the sun* (Halliday & Hasan 1976: 71ff.). Cohesive usage of the definite article involves in its clearest case the repetition of the item referred to (as in example 4.47), but often also use of a synonym, near-synonym, a hyponym (as in example 4.48) or even a meronym (Halliday & Hasan 1976: 72):

4.47 *The boy climbed on a tree. Then the boy looked into a hole.*

4.48 *The boy looked at the frog. Then the frog ran away. The kid was very sad.*

The indefinite article, on the other hand, indicates that an entity's referent is not recoverable from the preceding text. Any use of the indefinite article needs to be considered an exophoric demonstrative reference, which is non-cohesive and therefore excluded from the analysis. Instances of *all*, *some* or *both* as non-specific determiners were coded analogously.[44]

The classification of *there* was not always straightforward, since *there* occurs quite frequently in the data but not in its function as spatial adverb. Instead, it most often functions as existential *there* (example 4.49). In the data, *there* was only classified as a spatial adverb if it carried tonic prominence and allowed no combination with *here* (example 4.50; cf. Quirk et al. 1985: 1402ff.). Otherwise *there* was excluded from the analysis as being non-cohesive; in the analysis tables this is indicated by the presupposed element being coded as "test/ implicit".

4.49 1[there was a boy,] 2[he had a dog and a frog.] (C1-G4-2)

4.50 7[and the boy is looking in the glass,] 8[and the # frog is not there.] (C1-G1-16)

Comparative references occurred very infrequently. The ones identified in the data were *another*, which was split up for coding into indefinite article *an* and

[43] The NP-initial use of the definite article is cataphoric and thus non-cohesive (Halliday & Hasan 1976: 72).

[44] If *both* was used as a 'universal pronoun' (cf. Quirk et al. 1985: 381), it was coded in the same manner but with a subsequent nominal ellipsis.

comparative reference *other*, the comparative adverb *again*[45] and the adjective *next,* which was considered a comparative reference expressing difference. However, most occurrences of these comparative expressions were non-cohesive since they had an implicit (exophoric) reference point (cf. Quirk et al. 1985: 530, Halliday & Hasan 1976: 81, 324).

4.2.2.4 Substitution and ellipsis

Substitutions and ellipses occurred only very marginally in the data. Substitutions which occurred were either nominal (*one*) or verbal (*do*).[46] For the analysis, substitute *one* (example 4.51) needed to be differentiated from cardinal numeral *one* (example 4.52), which was coded as an instance of lexical cohesion, and *one* functioning as an indefinite article (example 4.53), which was excluded as being an exophoric demonstrative reference (cf. Hallliday & Hasan 1976: 98ff.). The following criteria were used: As opposed to the cardinal numeral, substitute *one* does not contrast with other numerals, it can have either singular or plural form (*one-ones*) and the substituted item cannot be added (example 4.54). The indefinite article *one*, on the other hand, is often accompanied by an ellipsis (example 4.52). At the same time, the substitute's plural form (*ones*) contrasts with the plural form of the indefinite article (*some*). Most importantly, however, substitute *one* is always accompanied by some contrast to the lexical item substituted (e.g. *another* in example 4.51); the expression encoding this contrast is phonologically salient, while substitute *one* is not (cf. Halliday & Hasan 1976: 95ff.). In this it resembles the indefinite article but contrasts with the numeral, which is always phonologically salient.

4.51 51[and there's sit his frog with another one.] (C5-G4-19)

4.52 79[and then … very much # frog childs ehm jumped out of # the grass.] # 80[and one # was jumping to the boy.] (C5-G4-21)

4.53 1[one night, a little boy had # a frog under his glass.] (C5-G4-19)

4.54 *Which toy? *The red one toy.*

Categories of ellipses encountered were nominal, verbal, and clausal. Any omission of (proper) noun, pronoun or (coordinated) noun phrase(s) was classified as nominal ellipsis and coded, while omissions of the verb or unified predicates were coded as being instances of verbal ellipsis (example 4.55; cf. chapter 4.2.2.1).

[45] Analogously to *more.*
[46] There were no instances of clausal substitution (*so, not*).

Ellipses involving a combination of nominal and verbal element(s), e.g. a verb phrase including a prepositional phrase (example 4.56), were classified as clausal and thus also coded. Other omissions, e.g. of {that} or {which}, were not considered.

4.55 65[then the deer be#guns <began> to run,] 66[and the dog in front of the deer.] (C1-G4-7)

4.56 17[then the dog fall out # the # window,] 18[and the boy also.] (C8-G1-1)

4.2.2.5 Connectives

All interclausal connectives were coded. Formally, this includes conjunctions (e.g. *and*), adverbs (e.g. *then*), prepositional phrases (e.g. *in the morning*), complex items (*one day*) or a combination of these forms. Semantically, Halliday and Hasan's (1976: 226) classification was used and the coding thus included additive, temporal, causal and adversative connectives (Tab. 4.6).[47] Three categories were added, however, namely conditional *if*, which Halliday and Hasan include under the general heading of causal (1976: 259), the "neutral" factual connective *that*, which introduces a specifying complement clause (cf. Biber et al. 1999: 658), and the subordinator *with* (cf. Quirk et al. 1985: 705).

Combinations of *and* or *but* plus a subsequent connective were coded as one unit, e.g. *and/ but then* as a temporal connective, since *and* (and to a lesser extent *but*) often functions as a discourse marker indicating that more is to be said (cf. Renkema 2004: 168, Berman & Slobin 1994c: 178, Halliday & Hasan 1976: 261).

Connective	additive	*and, or, also, and/ but ..also/ too/ as well*[48]
	temporal	*then, now, once, first, and then/ now, (and/ but) in the morning, (and/ but) one day, next day, suddenly, when, as…*
	adversative	*but, while, except*
	causal	*because, cause, for, (and) so*
	conditional	*if*
	factual	*that*
	subordinator 'with'	*with (the dog {being} on his bed)*

Tab. 4.6 Subtypes of connectives included in the analysis

Any complex connective item such as *in the morning* (e.g. C1-G4-1) was first of all coded as one unit according to its function as a connective, e.g. the temporal connective *in the morning.* As a second step the lexical items forming part of it were

[47] A more detailed formal and semantic analysis should be conducted in the future, since this study only considers the overall category connective. Additionally, a detailed functional analysis would be desirable, since form and function cannot simply be equated (cf. ch. 2.4). Unfortunately, such analyses are outside the scope of the present study.

[48] *And… also, and.. too* etc. are considered parallel constructions of additive connectives and thus coded as one unit (Halliday & Hasan 1976: 246).

analyzed for lexical ties, e.g. the oppositional relation *morning-night*. This seems justified since the use of a complex item represents a conscious lexical choice in expressing the semantic relationship between clauses much more specifically than if a more general connective is used, compare e.g. *in the morning* with *then*.

4.2.2.6 Lexical cohesion

The intra- and interclausal lexical relations of nouns, lexical verbs and adjectives were coded; morphological variants were considered together with their base forms. If a lexical item had no relationship to any preceding element, it was excluded as being non-cohesive. Two large categories of lexical cohesion, as distinguished by Halliday and Hasan (1976), were included in the analysis, namely reiteration and collocation (1976; cf. ch. 2.4). In the following, I will describe these two categories and the way they were operationalized for the analysis. After that, some general coding problems and their solutions will be described.

The category of reiteration comprises ties through simple repetition, synonymy, near-synonymy and hyponymy. The instances of ties by synonymy and near-synonymy identified and coded are listed in table 4.7; included in this category are instantial synonyms, e.g. *frog- Bingo* or *dog- Bello* (C5-G4-3; cf. also Hasan 1984). The occurrences of hyponymy identified are listed in table 4.8. Only the relationship hypernym-hyponym was coded, co-hyponyms were included in the collocation category described below.

Synonyms	*look- peep; spring- jump; tiny- little- small; great- big; earth- ground;* **instantial:** *dog-Bello…*
Near-synonyms	*look- see; see- peep; look- watch; cry- scream- shout- yell- say- call; hear- listen; bowl-glass*

Tab. 4.7 Synonymy and near-synonymy included in the analysis

Hypernym	Hyponyms
animal	*dog, bees, owl, mouse, deer, hamster, bird*
pet	*dog, frog*
kid/ child	*boy*

Tab. 4.8 Hyponymy included in the analysis

The second large category of lexical cohesion distinguished by Halliday and Hasan, namely collocation, comprises lexical items tied by opposition, co-hyponymy,[49] (co-)meronymy and general co-occurrence (1976; cf. also ch. 2.4). The

[49] This is taken to include sequences such as cardinal numerals, which can be described as co-hyponyms of *number*.

oppositional relations identified and coded are shown in table 4.9; only one instance of a co-hyponymic sequence, namely cardinal numerals, was coded as such.

Oppositional relations	
jump/ run/ climb/ creep/ go/ hop/ fall/ spring/ swim/ fly[50]	stand/ sit / lie/ stop/ stay
stand	sit/ lie
cry/ scream/ shout/ yell/ say/ call	hear/ listen
take	give
wake up	sleep
morning/ day	night/ evening
come	go
moon	sun
big/ great/ large	little/ small/ tiny
happy	sad/ unhappy
girl	boy

Tab. 4.9 Oppositional relations included in the analysis. Note: Slashes indicate equivalent alternatives. No distinction was made between different types of opposition such as antonyms and converses.

All other ties through collocation, including all instances of (co-)meronymy,[51] were operationalized as the word fields listed in table 4.10, which incorporate lexical as well as psycholinguistic relations in the sense of belonging to a common schema ("general co-occurrence").

No.	Word field	Members
1	**Going to bed and getting up**	*night/ morning or day, sleep/ wake up, bed, evening; bed - dream; sleep - dream; night - dark; night - dream, moon, pyjama, sleep; moon/ sun - shine; night, light, moonlight, evening; morning - sun, breakfast;*
2	**Forest/ woods**	*wood, tree, twig, bushes, (tree) trunk, log, stem, leaves*
3	**Animals**	*dog, frog, bees, owl, mouse, deer, hamster*
4	**Verbs of motion**	*jump, run, climb, creep, go (in the sense of movement), hop, spring, crabble, fall, fly, swim*
5	**Verbs of "perception"**	*see, look, peep - hear, listen - smell*
6	**Frogs and their habitat**	*frog, pond, lake, water, sea; wet, swim, shower, pond, lake, water, sea;*
7	**Animals and their associated activities**	*fly - bees, flies, bird; owl - fly; skunk - stink; bees - honey, sting*
8	**Sounds, sound production and perception**	*noise, sound, shout, loud, quiet, cry, hear, say, scream, yell, make pst, sing, specific sounds such as a quak, to quack, bark,*
9	**Mental verbs**	*think, believe, know*
10	**Body parts of animate beings**	*head, nose, hand, face, arms, face, tongue, cheeks, tummy, feet, paws, antlers, horns, muzzle, skin, horns*
11	**Animals and the sounds they make**	*frog - quak; dog- bark;*
12	**Emotions and their expression**	*unfriendly/ friendly, nice, sad, unhappy/ happy, angry, laugh, afraid, smile, cry, kiss, like, love, surprised*
13	**House and furniture**	*house, window, windowsill, room, bed, chair, door, carpet*
14	**Family**	*baby, parent, father, mother, dad, married, man, wife, children, kid, baby, woman, girlfriend*
15	**Temporal expressions**[52]	*time, moment, night, morning, evening*

Tab. 4.10 Word fields included in the analysis. Note: More specific lexical relations, i.e. (near-)synonymy, hyponymy and oppositional relations, were coded with preference. Slashes indicate a relationship of opposition between the terms separated by them. Terms separated by commas are in a relationship of

[50] Corresponds to word field 4 "verbs of motion."

[51] (Co-)meronymy was not coded as a separate lexical relation, since meronymic relations are often difficult to delimit (cf. Cruse 2004, Löbner 2002). The same is true for (co-)hyponymy of verbs.

interchangeability, while separation by a semi-colon indicates that only the terms in this chain were considered interchangeable in coding.

Finally, some further decisions made in the coding of lexical ties need to be outlined. First of all, the analysis was limited to relations between individual lexical items. Relationships between phrases and lexical items were not coded, although e.g. an opposition between *could not come out anymore* and *free* (example 4.57) is in principle not different from an opposition such as *stuck* - *free*, which would have been coded.

4.57 12[but he could not come out anymore.] …17[the glass exploded,] 18[and the dog was free.] (C1-G4-10)

Secondly, lexical verbs similar to modals and aspectuals were excluded from the analysis, e.g. *let* (example 4.58) or *want,* since their function is to modify the meaning of the following lexical verb (cf. ch. 4.2.2.1).

4.58 23[and the deer get the boy] 24[and # let him fell in a pond.] (C1-G4-2)

Similarly, forms of *be* and lexical verb *do* were excluded due to their minimal semantic content (cf. Halliday & Hasan 1976: 125f.) and also *go* in fixed expressions such as *go to sleep*, since in contrast to *go to bed* or *go home* no interpretation of movement in space is possible. *Have* was only coded where it could be replaced with *possess* or *own.*

Compounds were treated as one unit where a relationship of hyponymy was identifiable as the most immediate lexical relation (e.g. *mouse hole* - *hole*). Otherwise the most immediate relationship of one of their parts was coded. This allowed for ties through repetition (e.g. *mouse hole* - *mouse*) or membership in a lexical field (e.g. *mouse hole* - *dog* as common members of the lexical field 'animal').

Finally, the story's title *Frog where are you?* and its variants, e.g. *where are you, frog?*, were split up into their individual elements – except if they were used with the actual function of a title at the beginning of a story as in example 4.59.

4.59 1[Frog, where are you?] # 2[once upon a time…] (C1-G4-3)

In the latter case they were excluded as being an exophoric reference to the picturebook (non-cohesive). Subsequent mentions, however, were always coded according to their individual elements, even if an exophoric use referring to the title could not be excluded for any of the occurrences.

[52] Lexical field 15 was created for lexical cohesion not covered by the categories oppositional relation or lexical field 1.

4.2.3 Statistical methods

In addition to the descriptive statistics, several inferential statistical tests were performed. These will be described very briefly in the following in order to facilitate interpretation in the results section.

General linear model analyses of variance (ANOVA) were conducted in order to establish main and interaction effects of the independent variables grade, sex, and L2 experience on a dependent variable, namely participants' total number of narrative components and cohesive density. ANOVAs serve to determine significant differences between several means on the basis of within-group and between-group variances (Mackey & Gass 2005: 274).[53] Results are presented in the form of the *F* value, which represents the ratio of between-groups over within-group variance (Hatch & Lazaraton 1991: 315) and which must be larger than 1 to indicate at least some difference among groups (ibid. 321). A 2x2x3 design over all participants was carried out with grade (first, fourth), sex (female, male) and L2 preschool experience (mono, bili, MB, cf. ch. 4.1.2) as between-subjects factors. Further analyses of variance – with reduced designs e.g. excluding statistical outliers – were carried out where necessary; these are described in the corresponding chapters. Partial eta squared (represented as η_p^2) is presented as a measure for effect size, i.e. as an indicator for the degree of association, if F-values were statistically significant.[54]

Independent samples t-tests (reported as 't') were conducted to test the statistical significance of differences between means where the analysis of variance showed significant interaction effects. Based on the degrees of freedom t-tests allow to determine whether the means of two groups are statistically different or not (e.g. Rietveld & van Hout 1993: 15, Hatch & Lazaraton 1991: 249ff., Woods et al. 1986: 176ff.).[55] An analysis of variance may have found no overall main effect of experience group for one of the variables tested, for example, but an interaction

[53] ANOVAs cannot be conducted, however, when the underlying assumptions of homogeneity of variances is violated (cf. Field 2009:436). If this was the case, it is reported in the respective chapters and independent samples t-tests were carried out instead.

[54] Any measure for effect size or degree of association indicates "how much of the variability in the data can be accounted for by the independent variable" (Hatch & Lazaraton 1991: 266). Thus, (partial) eta squared can be interpreted "as the proportion of the total variability of the dependent variable which is explained by the variation in the independent variable" (Porte 2002: 235). An eta square value of 0,45, for example, means that 45% of the variability in the sample can be explained by the independent variable under investigation (cf. Hatch & Lazaraton 1991: 266). At the same time it indicates that another 55% of the variability is left unexplained.

[55] SPSS first of all tests whether the two groups have equal or unequal variances (Levene test). If the two groups have equal variances, a t-test with (N-2) degrees of freedom is conducted, while in the case of unequal variances a more conservative t-test with a lower degree of freedom is applied.

effect of grade and experience. In a second step t-tests would then have been conducted for grade 1 and 4 separately in order to test the statistical significance of differences between experience groups in each grade.[56] Additionally, eta square (represented as η^2) was calculated as a measure for the effect size of the t-test result (e.g. Porte 2002: 121; cf. also η_p^2 above).

Pearson chi square tests were used to test the statistical significance of observed differences in frequency variables[57] (e.g. Hatch & Lazaraton 1991: 394, Woods et al. 1986: 139), namely differences in the realization of individual narrative components possibly attributable to grade, sex and experience group. Chi square tests indicate whether two variables are statistically independent[58], i.e. "different" from each other, or on the other hand associated, i.e. "related", by comparing observed frequencies with the ones to be expected in case of independence (e.g. Porte 2002: 136, Hatch & Lazaraton 1991: 393ff., Woods et al. 1986: 139ff.). The chi square value (reported as χ^2) thus indicates the amount of difference between expected and observed frequencies. The higher the χ^2 value the more likely this difference is statistically significant, i.e. the more likely it is that no relationship exists between the variables tested. Additionally, phi (reported as Φ) is presented as an indicator of strength of association for chi square tests.[59]

In case of expected cell frequencies below five as well as in case of a strongly unequal sampling distribution Fisher's exact test was used to determine significance.[60] Fisher's exact test follows the same logic as chi square, but it also accommodates small and unequally distributed sample sizes (cf. Hatch & Lazaraton 1991: 409). The significance level of differences tested with Fisher's exact test will be reported as 'Fisher's p' in the results section.

---

[56] The MB group was excluded from this step due to the small number of participants (N=3 in grade 1; N=4 in grade 4).

[57] Variables measuring how often a variable is present in the data ('nominal variables') as opposed to score data ('ordinal' and 'interval variables'), which shows how much of a variable is present (Hatch & Lazaraton 1991: 62). Nominal variables are categorical variables, i.e. coded dichotomously using 1 and 0 to represent e.g. "realized" and "not realized".

[58] That is, the unrelatedness of a dependent (e.g. number of narrative components) and an independent variable (e.g. L2 experience).

[59] While the value of chi square only indicates whether there is a relationship between variables, phi serves to determine how strong this relationship is in a 2x2 design (Hatch & Lazaraton 1991: 415f.; cf. also the description). Phi follows the same logic as (partial) eta squared described above: A phi value of 0,28, for example, can be translated as the independent variable (e.g. grade) accounting for 28% of the variability of the dependent one (e.g. number of narrative components), which would be considered a fairly strong relationship.

[60] Where expected cell frequencies dropped below one, however, the chi square results were discarded because in such cases the test cannot be considered reliable anymore (cf. SPSS 15.0 Manual) even if additionally Fisher's exact test is applied to determine significance. Thus, no statistical test was applied in such cases.

A Mann-Whitney U-test (reported as 'U') was used to compare score data[61] from two groups if the distribution of the data was not at least approximately normal (cf. Hatch & Lazaraton 1991: 270ff.), e.g. for a newly created index of global narrative structure (cf. ch. 5.3). This statistical test indicates whether there is a significant difference in ranks between two groups. Since Mann-Whitney U-test is a non-parametric test based on the median, normal distribution of the data is not required, which distinguishes this test from its more commonly used parametric counterpart, the two-sample t-test. In addition, eta square (represented as η^2) was calculated as a measure of strength of association (cf. Hatch & Lazaraton 1991: 279).

Cronbachs alpha, a coefficient of consistency, was used to assess test reliability for a newly created narrative index (cf. ch. 5.3). After the construction of a scale or an index according to validity considerations Cronbach's alpha, which is based on inter-item correlation, can be used to determine the reliability of test items in terms of internal consistency (e.g. Mackey & Gass 2005: 128ff., Schmitt 1996, Cortina 1993). Thus, the alpha coefficient indicates on a range from 0 to 1[62] to what degree a combination of test items is able to measure a single construct (cf. Eckstein 2008: 293ff., Rietveld & van Hout 1993: 203). Individual test items can be included or eliminated based on considerations about their contribution to the overall alpha score. Even though no standard critical alpha coefficient exists, an alpha value around 0,7 or higher is generally considered acceptable, especially if a small number of items is tested (Eckstein 2008: 293ff., Schmitt 1996, Cortina 1993).

Pearson correlation analyses[63] were conducted to investigate the relationship between narrative coherence and cohesion. Correlation indicates whether two variables are linearly related and the correlation coefficient[64] additionally indicates how strong this relationship is (cf. Hatch & Lazaraton 1991: 427). This linear relationship is indicated by a correlation coefficient ('r') between -1 and 1. A coefficient of 0 indicates that there is no relationship, while a coefficient between -1 and 0 indicates a negative correlation, which means that the two variables perform in opposite ways, i.e. one variable has a low value if the other one has a high value and the reverse. A coefficient between 0 and 1 consequently indicates a positive

[61] Mann-Whitney U-test can be applied to score data, while chi square is used for nominal or frequency data, e.g. *Yes* vs. *No* answers (cf. above). Mann-Whitney U-test is generally considered the non-parametric alternative to a t-test.
[62] If test items are positively correlated, otherwise no lower limit exists (Eckstein 2008: 293).
[63] In SPSS this includes point biserial correlations, i.e. correlations involving a dichotomous variable. Pearson correlation is used if at least one of the variables is ordinal.
[64] In the case of Pearson correlation the coefficient is symbolized by *r.*

correlation in the sense that both variables perform the same having either both high or both low values. The closer the correlation coefficient is to +/-1, the stronger is the relationship investigated (Hatch & Lazaraton 1991: 433). In addition the correlation coefficient r^2 was calculated as a measure for the strength of the relationship.[65]

4.3 Addressing the dangers of the comparative fallacy

Any study on L2 acquisition has to take care to avoid what Bley-Vroman (1983) calls the 'comparative fallacy', which addresses the dangers of comparing learner languages with the respective target languages (i.e. learner language systems with L1 language systems), especially with respect to defining analytical categories through target-language standards. Bley-Vroman states that

> [f]or example, any study which classifies interlanguage (IL) data according to a target language (TL) scheme or depends on the notion of obligatory context or binary choice will likely fail to illuminate the structure of the IL. Of course any study which uses a target language scheme to preselect data for investigation (such as a study which begins with a corpus of errors) is even more liable to obscure the phenomenon under investigation. (1983: 15)

At the same time Bley-Vroman's justified methodological criticism cannot obscure the fact that "[g]iven that SLA is the study of how people acquire a second language, any SLA study implicitly has a built-in notion of interlanguage with a target language lurking in the background" (Park 2004: 3). Consequently, other researchers recognize a target-language perspective as one of several possible and legitimate perspectives on studying learner language (e.g. Cook 1999: 190, Ellis 1994). The reader is referred to Kersten (2009: 60ff.) for a very recent discussion of different possible approaches to studying learner language (interlanguage, L1 perspective, L2 perspective). Only a combination of all of these perspectives can, however, come even close to a full description of a system as complex and variable as that of learner language(s).

Since any research perspective carries its own methodological problems (e.g. Kersten 2009: 60ff., Foster-Cohen 2001, Lakshmanan & Selinker 2001), however, I will opt for a "pragmatic" stance with respect to my analysis. That is, I will not contribute further to the debate which perspectives are (not) appropriate and under which circumstances. Instead, I will describe very briefly how my analysis aimed to at

[65] R^2 indicates how much of the variance of one variable is explained by the other, i.e. the degree to which the two variables provide the same information (Hatch & Lazaraton 1991: 441).

least reduce the danger of falling prey to the comparative fallacy and the reader is asked to form his or her own opinion as to how far this was successful.

Just as the theoretical basis put forward for the analysis of coherence and cohesion, my method of analysis could best be described in terms of a combination of top-down and bottom-up processes: On the one hand, both coherence and cohesion analysis make use of certain general categories, namely the narrative components described in ch. 4.2.1 and the categories of cohesion identified by Halliday and Hasan (1976) (cf. ch. 4.2.2.2). All of these categories have proved useful in previous studies on L1 and/or L2 learners (cf. ch. 3). The first step in setting up the analysis grid for both coherence and cohesion was, however, to systematically evaluate how the study's participants realized these general categories linguistically, i.e. which options they used in their L2 productions. For both coherence and cohesion not only target-like options were coded.

For the analysis of coherence the overall criterion applied was communicative success, i.e. whether the interlocutor – who did pretend, however, to be a native speaker of English – would be able to comprehend the speaker and fill each narrative component of the story schema with content. At the same time only a minimum of target-likeness was required, especially with respect to syntax, morphology and phonology, i.e. the interlocutor's purported "comprehension threshold" was set as low as possible. This process was aided by the researcher, i.e. the author, being an advanced, but non-native speaker of English herself (cf. Lakshmanan & Selinker 2001: 415).

For the analysis of cohesion a similar path was followed. That is, the general categories identified by Halliday and Hasan (1976) were applied to the data but their linguistic realization was not measured exclusively by L1 standards: One child (C1-G1-5), for example, produced *and on the day*, which was coded as a complex connective, even if an (adult) native speaker would most probably not have produced such an expression in the same linguistic context. Other children, for example, systematically pronounced the definite article *the* as /se/ but these non-target-like forms were of course still coded as definite articles. At the same time all instances of realizations of the cohesion categories were classified, regardless of whether these would, for example, constitute an over- or underuse when compared to L1 speakers.

Since any linguistic study is to a certain degree an artefact of its methodology, the most important precaution (not only against the comparative fallacy) was to give

a detailed description of my method of analysis in the previous sections of the present chapter. This was done to ensure, first of all, that a replication of the study is possible. Secondly, this detailed description also serves to allow readers to form their own opinion about any possible bias in results attributable to the method of analysis – such as to which degree my study is subject to the comparative fallacy.

5 The development of coherence: Results

5.1 Total number of narrative components

This section is aimed at answering the question (a) how coherent participants' stories were as measured by the number of narrative components realized in first and fourth grade and (b) whether there were any quantitative differences attributable to grade, sex or experience. To this end I will describe first of all the observed results for the overall number of narrative components in first and fourth grade as well as the results obtained for the two background variables included in the study, namely participants' sex and L2 preschool experience. After that I will present the corresponding statistical results.

As stated in ch. 4.1.2, possible differences between the longitudinal and cross-sectional data sets were explored for all overall measures of coherence (and cohesion) as well as for randomly selected individual variables. Since no significant differences were found, no further distinction will be made between cohorts in the discussion of the results.

5.1.1 Total number of components: Observed results

The narrative components actually realized by participants corresponded to a very large extent to the 14 narrative components identified in ch. 4.2.1, namely

- setting
- initiating event
- simple reaction
- goal 1 and 2
- attempts 1 to 7
- consequence
- ending.

The only exceptions from the pre-identified components were produced by children C8-G1-13 and C8-G1-20, who realized two attempts in addition to the ones previously identified. These results speak for the validity of the approach.

5.1.1.1 Overall results

The results obtained for the total number of narrative components[1] in first and fourth grade is shown in Fig. 5.1 together with its increase from first to fourth grade; the corresponding descriptive statistics are given in Tab. 5.1.

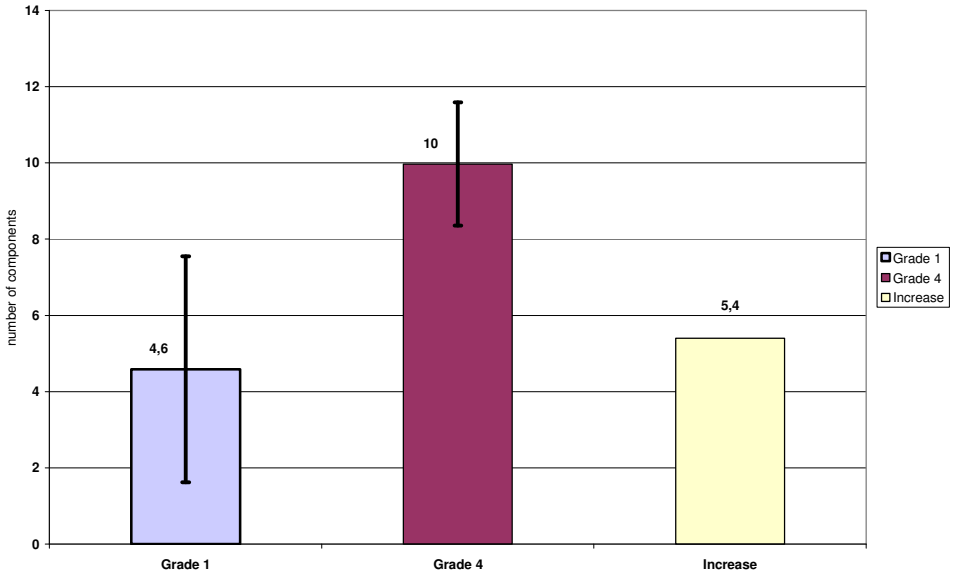

Fig. 5.1 Mean number of narrative components by grade as well as increase from first to fourth grade

	Total N	Mean	Standard Deviation	Median[2]	Mode[3]	Minimum	Maximum
Grade 1	34	4,6	3,0	4,5	1	1	10
Grade 4	32	10	1,6	10	10	6	12
Total	66	7,2	3,6	8	10	1	12

Tab. 5.1 Descriptive statistics total number of narrative components by grade

All first graders were able to realize at least one component and on average they produced 4,6 narrative components. None of the first graders, however, produced more than 10 of the 14 components tested. At the same time the interindividual differences are very large in grade 1, which is evident from a standard

[1] Only the components previously identified (cf. ch. 4.2.1) were considered.
[2] The median indicates the center of a distribution (Hatch & Lazaraton 1991:161). That is, 50% of all scores lie above and 50% below the median.
[3] The mode indicates the most frequent score. More than one mode can exist.

deviation of three components, i.e. a variability of 65% in relation the mean (in the following 'variability ratio'), and a range of 9 from minimum (1 component) to maximum score (10). This strong heterogeneity of results is confirmed by a look at Fig. 5.2, which presents the distribution of the number of components realized in grade 1 and 4; the corresponding descriptive statistics are given in Tab. 5.2.

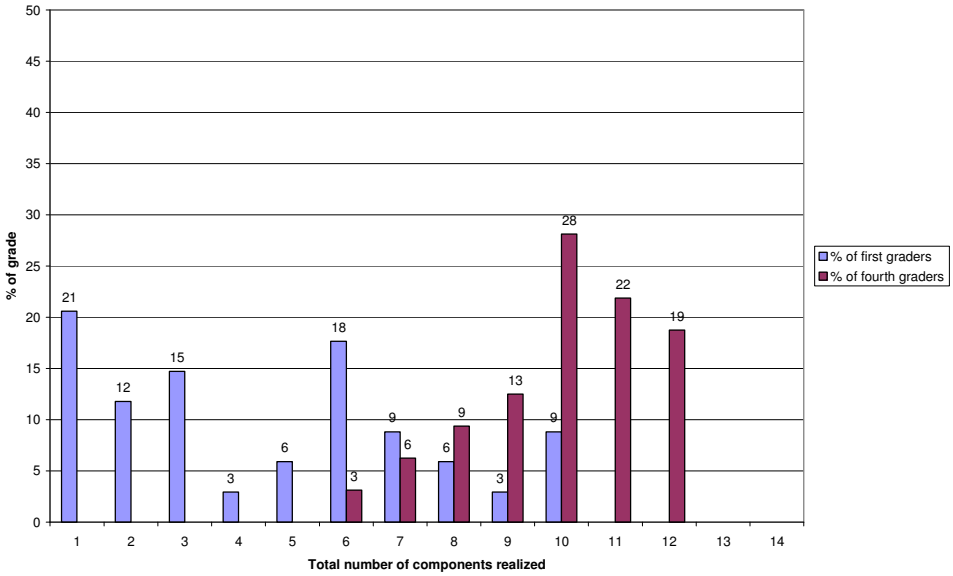

Fig. 5.2 Distribution of the total number of component scores by grade

		Grade 1		Grade 4	
		N	% within grade	N	% within grade
Total number of narrative components	1	7	21%	0	0%
	2	4	12%	0	0%
	3	5	15%	0	0%
	4	1	3%	0	0%
	5	2	6%	0	0%
	6	6	18%	1	3%
	7	3	9%	2	6%
	8	2	6%	3	9%
	9	1	3%	4	13%
	10	3	9%	9	28%
	11	0	0%	7	22%
	12	0	0%	6	19%
Total		34	100%	32	100%

Tab. 5.2 Distribution of the total number of component scores by grade

The distribution shows that in terms of frequency first graders most often realized 1 (21%, i.e. 7 out of 34 first graders) or 6 components (18%, 6 of 34).

Slightly less often participants produced 3 (15%, 5 of 34) and 2 components (12%, 4 of 34). All other components were realized by less than 10% of first graders.[4] At the same time the values below the mean, i.e. less than 4,6, are more evenly distributed: The realization of 1, 2 or 3 components already accounts for 48% of the results, while values higher than the mean, i.e. from 4,6 to 10 components, are realized mainly by a strong clustering of scores at 6 components (18%). The highest score (10 components) is achieved by only 9% of first graders (3 of 34, namely C1-G1-16, C8-G1-13 and -21). One could also say that the distribution shows a division of first graders into minimal (1 to 3 components) and proficient storytellers (6 or more components).

Fourth grade participants realized an average of 10 and minimally 6 components. That is, the mean more than doubles from first to fourth grade – from 4,6 to 10 components or from 33% to 71% of the 14 components tested. At the same time the minimum score increases very strongly (from 1 to 6 components). However, even in fourth grade the highest score is 12 so that none of the participants in either grade produced all 14 narrative components (Tab. 5.2, Fig. 5.2). Additionally, participants' performance becomes far more homogeneous from first to fourth grade (Fig. 5.1, Tab. 5.1). Thus, their interindividual differences decrease from a variability ratio of 65% of the mean in first to only 16% in fourth grade. Similarly, the range of scores from minimum to maximum decreases from 9 to 6 (Tab. 5.1). This development is reflected in Fig. 5.2, which shows low frequencies below 10 components and a clustering of observations above that.

Taking a closer look at the distribution in grade 1 and 4 (Fig. 5.2) an overlap of scores can be observed between 6 and 10 components (fourth grade lowest to first grade highest score). This overlap includes 44% of all first graders (15 of 34) and 59% of fourth graders (19 of 32). Scores below 6, on the other hand, are exclusive to first and scores above 10 to fourth graders. Judging from the frequency of realizations values between 1 and 6 components seem to be most characteristic for grade 1, however, since together they account for 75% of all observations in first grade. In grade 4, on the other hand, values of 10 components or more seem to be most typical, since they account for 69% of all observations, while observations below 10 components are more spread out and account for only about one third of fourth graders.

[4] However, in absolute numbers the differences are relatively small. The difference between 9% and 12%, for example, corresponds to one child.

To summarize, a comparison of the overall number of components by grade showed that all participants were able to produce at least one component. At the same time a dramatic increase in the number of components was found from first to fourth grade paired with a strong decrease in interindividual variation. A score of 6 components or less was found to be most typical for first graders, while a score of 10 components or more was identified as most typical for fourth graders. Almost half of all first graders, however, were found to produce a number of components which could also have been produced by a fourth grader.

5.1.1.2 Total number of components by sex

Fig. 5.3 and Tab. 5.3 show the results obtained in first and fourth grade as a function of participants' sex.

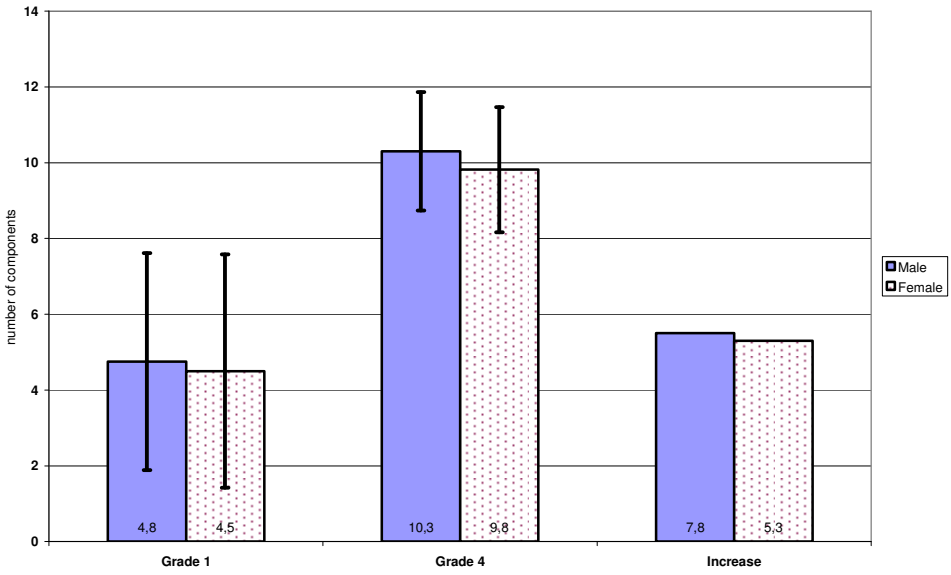

Fig. 5.3 Mean number of narrative components by sex and grade as well as increase from first to fourth grade

		Total N	Mean	Standard Deviation	Median	Mode	Minimum	Maximum
Grade 1	Male	12	4,8	2,9	5	6	1	10
	Female	22	4,5	3,1	4	1	1	10
	Total	34	4,6	3	4,5	1	1	10
Grade 4	Male	10	10,3	1,6	10,5	12	8	12
	Female	22	9,8	1,7	10	10	6	12
	Total	32	10	1,6	10	10	6	12

Tab. 5.3 Descriptive statistics total number of components by sex and grade

Only slight descriptive differences exist between male and female participants in either grade in terms of the mean number of components or the interindividual differences as measured by standard deviation and range. Female first graders produce an average of 4,5 and male first graders a slightly higher mean of 4,8 components. At the same time the ratio of standard deviation to mean is 69% for female first graders and 60% for males – slightly above and below, respectively, of the average variability ratio of 65% in first grade.

Both sexes follow the general trend of strongly increasing their performance. Consequently, their fourth grade results are again very similar with males producing a mean number of 10,3 and female participants a mean number of 9,8 components. At the same time both groups' interindividual variation decreases and they become even more similar in this respect. Thus, female fourth graders' variability ratio is 17% and male fourth graders' 16%.

Still, male participants seem to perform slightly better in both grades: In first grade, the median of male participants is 5 components, indicating that 50% of males realized 5 or more components, while for female participants the median lies at 4 components (Tab. 5.3). Similarly, the mode, i.e. the number of components realized most frequently, is 6 components for male and 1 for female first graders; this is reflected in Fig. 5.4.

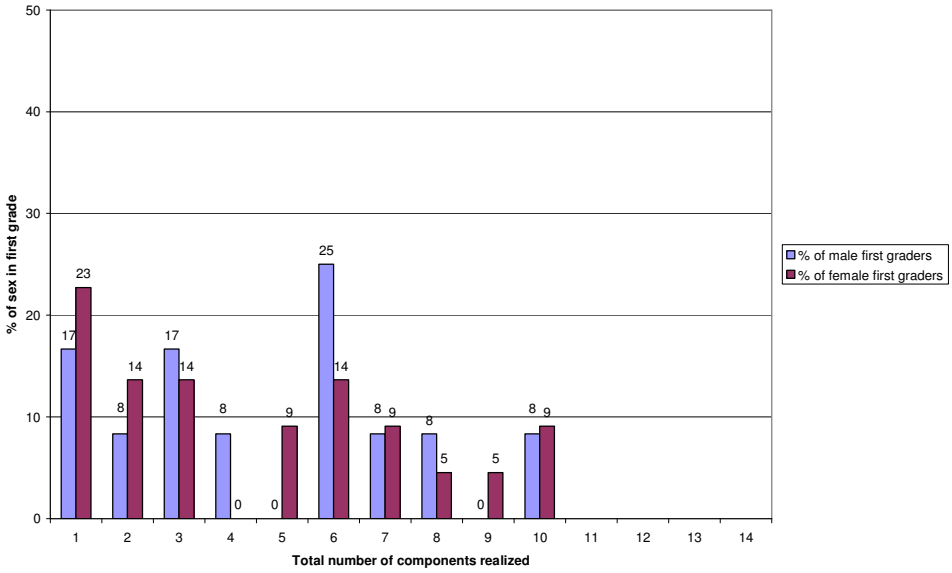

Fig. 5.4 Distribution of total number of narrative component scores by sex in first grade

		Grade 1				Grade 4			
		Male		Female		Male		Female	
		N	% within sex	N	% within sex	N	% within sex	N	% within sex
Total number of narrative components	1	2	17%	5	23%	0	0%	0	0%
	2	1	8%	3	14%	0	0%	0	0%
	3	2	17%	3	14%	0	0%	0	0%
	4	1	8%	0	0%	0	0%	0	0%
	5	0	0%	2	9%	0	0%	0	0%
	6	3	25%	3	14%	0	0%	1	5%
	7	1	8%	2	9%	0	0%	2	9%
	8	1	8%	1	5%	2	20%	1	5%
	9	0	0%	1	5%	1	10%	3	14%
	10	1	8%	2	9%	2	20%	7	32%
	11	0	0%	0	0%	2	20%	5	23%
	12	0	0%	0	0%	3	30%	3	14%
Total		12	100%	22	100%	10	100%	22	100%

Tab. 5.4 Distribution of the total number of component scores by sex and grade

In fourth grade, no difference in median can be observed but still a difference in modes with 12 components for male fourth graders (30%) but 10 components for female fourth graders (32%) (Tab. 5.3). This is also reflected in Fig. 5.5, which shows the distribution of male and female fourth grade scores.

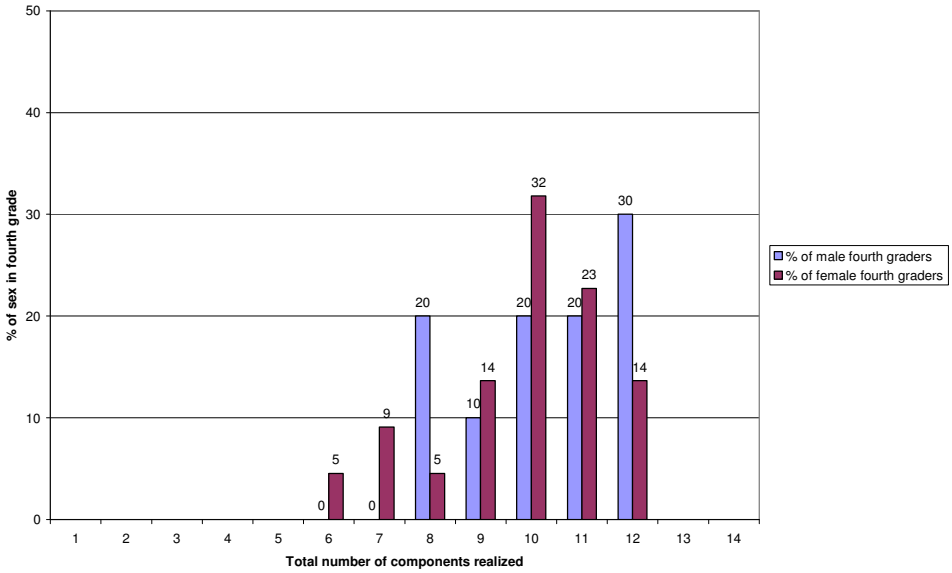

Fig. 5.5 Distribution of the total number of components by sex in fourth grade

Additionally, the distribution shows that 50% of male fourth graders (5 of 10) realize 11 or more components as opposed to only 36% of female participants (8 of 22). At the same time the lowest scores, namely 6 (5%, C5-G4-20) and 7 components (9%, 2 of 22), were produced exclusively by female fourth graders. That is, there seems to be a slight male advantage in first as well as fourth grade. However, this finding cannot be considered solid due to the likelihood of a group size effect.[5]

In regard to male as well as female results in first as compared to fourth grade both groups show some overlap. Male participants' first and fourth grade distribution of the total number of component scores is depicted in Fig. 5.6 and the corresponding female distribution in Fig. 5.7.

[5] In absolute numbers, for example, three of the male (25%) as well as three of the female participants (14%) realize six components.

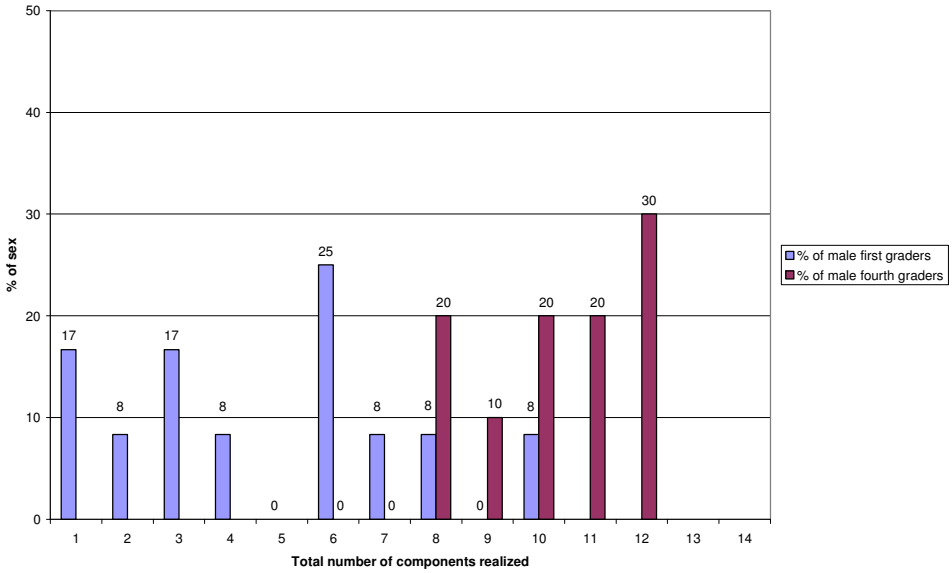

Fig. 5.6 Distribution of the total number of components by grade for male participants

Male first and fourth grade results overlap in the range of 8 to 10 components (male fourth grade minimum to first grade maximum). Only 16% of male first graders, however, realized between 8 or 10 components, which corresponds to one male first grader scoring 8 and another one 10 components. These latter two male first graders seem to be precocious in their storytelling abilities in comparison to their peers, since all remaining male first graders produced fewer than 8 components.

50% of male fourth graders (5 of 10), on the other hand, realized between 8 and 10 components and the other 50% more than 10. This means that scores below 8 components are exclusive to first and scores above 10 exclusive to fourth graders, which indicates that some male first graders do already perform like fourth graders but the great majority of them does not. Thus, the distribution of scores clearly confirms a quantitative difference between first and fourth grade results.

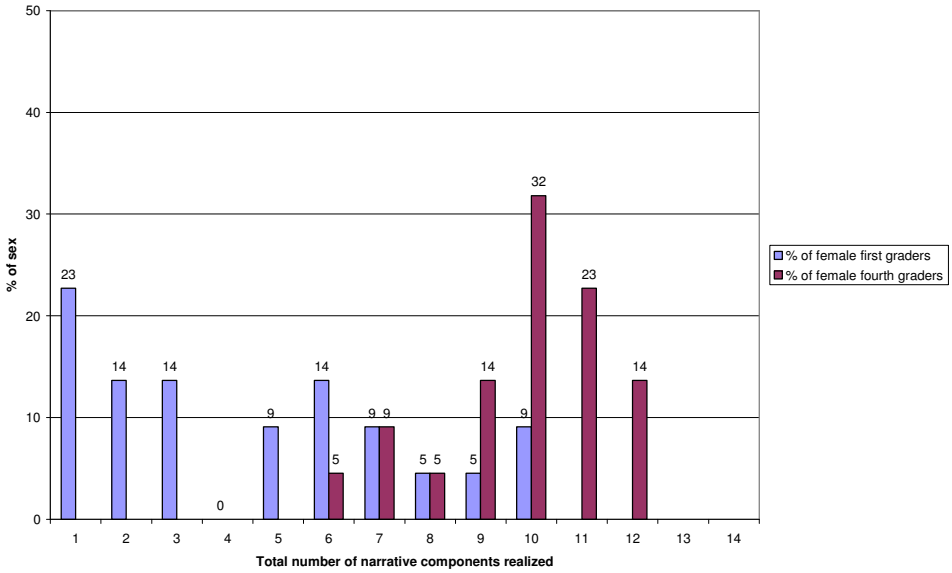

Fig. 5.7 Distribution of the total number of components by grade for female participants

With respect to female participants the difference between first and fourth grade results is a little less pronounced. Female results overlap in a somewhat larger range, namely between 6 and 10 components (female fourth grade minimum to first grade maximum). 41% of female first graders (9 of 22) produced a total number of components within this range, while the majority scored below (59%, 13 of 22). At the same time the only score within the overlapping range which was produced by more than just one or two female first graders was 6 components. Scores of 7 or more components were realized by only one (5%) or two female first graders (9%). The majority of female fourth graders (64%, 14 of 22), on the other hand, produced results within the overlapping range; 36% of female fourth graders (8 of 22) scored higher. However, the three lowest fourth grade scores, namely 6, 7 and 8 components, are produced by 5% to 9% of fourth graders, which correspond again to merely one and two participants, respectively. Thus, a result of 6 or fewer components is more typical for female first graders, while 9 or more components seem to be more typical for female fourth graders.

To sum up, male and female results closely resemble each other as well as the overall results described earlier, i.e. there are only slight differences between sexes in first and fourth grade. At the same time the mean for both groups strongly increases from first to fourth grade and their results become more homogeneous. Male and female distributions both indicate an overlap between first and fourth grade scores with the difference between the two grades being more pronounced for male

participants. The respective distributions also indicated that scores of 6 components or less are most typical for female and 8 or less for male first graders. In fourth grade, scores of 9 or more components were found most typical for female and scores of 10 or more for male participants.

5.1.1.3 Total number of components by experience group

The total number of components as a function of L2 preschool experience and grade is shown in Fig. 5.8; the corresponding descriptive statistics are given in Tab. 5.5. In the following I will first of all describe and compare the results obtained for mono and bili group; the MB group will be considered separately due to its small size and thus severely limited representativeness.

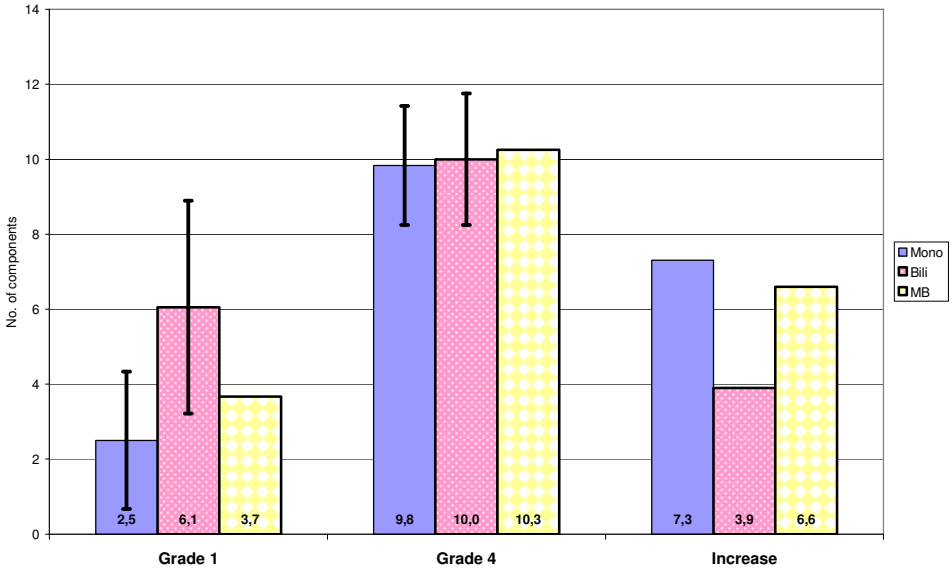

Fig. 5.8 Mean number of components by grade and experience as well as increase from first to fourth grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children who attended a monolingual group in a preschool offering also bilingual groups.

		Total N	Mean	Standard Deviation	Median	Mode	Minimum	Maximum
Grade 1	Mono	12	2,5	1,8	2	1	1	6
	Bili	19	6,1	2,8	6	3	1	10
	MB	3	3,7	-	-	1	1	6
	Total	34	4,6	3	4,5	1	1	10
Grade 4	Mono	12	9,8	1,6	10	10	7	12
	Bili	16	10	1,8	10	12	6	12
	MB	4	10,3	-	-	11	8	11
	Total	32	10	1,6	10	10	6	12

Tab. 5.5 Descriptive statistics total number of narrative components by experience group and grade. Note: More than one mode exists. 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children who attended a monolingual group in a preschool offering also bilingual groups.

In first grade the bilingual experience group clearly outperforms the monolingual one: On average bilis realize more than twice as many components (6,1 vs. 2,5). A similar advantage is evident in the distribution of scores (cf. also the respective modes), which is depicted in Fig. 5.9; the corresponding values are given in Tab. 5.6. Almost half of all first graders with monolingual preschool experience (42%) realize only one component, while those with bilingual preschool experience most often realize 3 (21%) or even 6 components (21%).

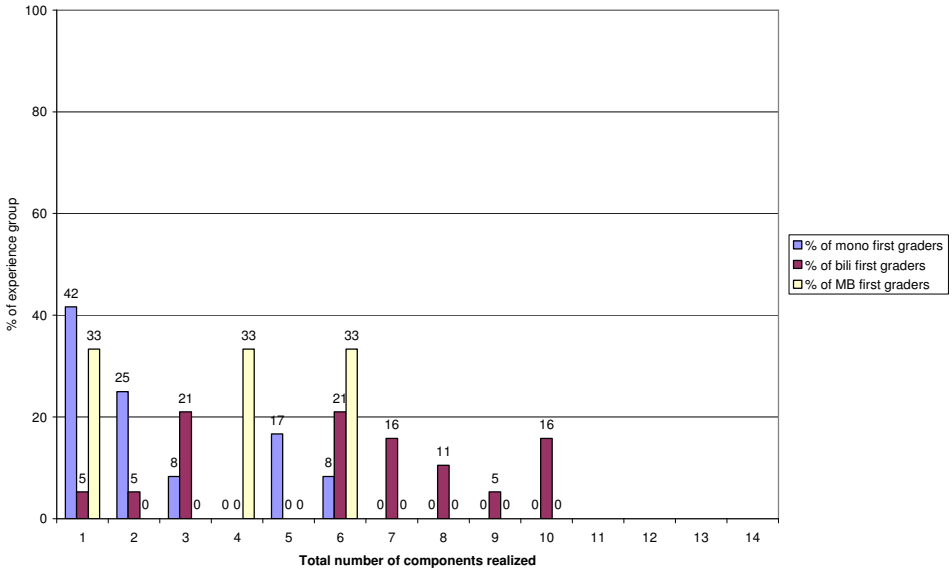

Fig. 5.9 Distribution of the total number of components by experience group in first grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children who attended a monolingual group in a preschool offering also bilingual groups.

		Grade 1						Grade 4					
		Mono		Bili		MB		Mono		Bili		MB	
		N	% within exp. group	N	% within exp. group	N	% within exp. group	N	% within exp. group	N	% within exp. group	N	% within exp. group
Total number of narrative components	1	5	42%	1	5%	1	33%	0	0%	0	0%	0	0%
	2	3	25%	1	5%	0	0%	0	0%	0	0%	0	0%
	3	1	8%	4	21%	0	0%	0	0%	0	0%	0	0%
	4	0	0%	0	0%	1	33%	0	0%	0	0%	0	0%
	5	2	17%	0	0%	0	0%	0	0%	0	0%	0	0%
	6	1	8%	4	21%	1	33%	0	0%	1	6%	0	0%
	7	0	0%	3	16%	0	0%	2	17%	0	0%	0	0%
	8	0	0%	2	11%	0	0%	0	0%	2	13%	1	25%
	9	0	0%	1	5%	0	0%	1	8%	3	19%	0	0%
	10	0	0%	3	16%	0	0%	6	50%	3	19%	0	0%
	11	0	0%	0	0%	0	0	1	8%	3	19%	3	75%
	12	0	0%	0	0%	0	0	2	17%	4	25%	0	0%
Total		12	100%	19	100%	3	100%	12	100%	16	100%	4	100%

Tab. 5.6 Distribution of the total number of components by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children who attended a monolingual group in a preschool offering also bilingual groups.

The frequency distribution also shows an almost steady decline in the percentage of mono first graders realizing more than one component: 75% of the mono first graders (9 out of 12) realize one, two or three components. Only 25% (3 children), however, produce more components, namely five or six. The highest mono score of six components is reached by only 8% of them, i.e. one child (C1-G1-10). At the same time, none of the mono first graders realizes 4 components, and this seems to split them into a low and a high performance subgroup. Thus, a score of three components or less seems to be more typical for first graders without prior L2 preschool experience.

The bilingual experience group has a larger range of scores – from a minimum of 1 to a maximum of 10 components – and its scores are more evenly distributed across that range. This is especially evident in the absolute frequencies (cf. Tab. 5.6). At the same time none of the bili first graders realizes four or five components and this seems to also split the bili group up into a low and a high performance subgroup: 32% of the bili children (6 of 19) realize one, two or three components, while 68% (13 of 19) realize six or more. Thus, a score of six or more components

may be even more typical for first graders with L2 preschool experience and a score of three components or less for first graders from monolingual preschool groups.

The interindividual differences as measured by the standard deviation (Tab. 5.5) confirm the more even distribution in the bili group: The ratio of the standard deviation to the mean is much larger in the mono (72%) than in the bili group (45%), which means that the results achieved by the mono group are more heterogeneous, even if nominally the standard deviation in the bili group is larger. However, both groups have a comparatively large interindividual variation in first grade.

The MB group, i.e. children from a monolingual group of the bilingual preschool, seems to take an intermediate position between mono and bili first graders (Tab. 5.5). However, since this group is so small (N=3) and at the same time very heterogeneous in the distribution of scores (cf. Tab. 5.6 & Fig. 5.9), no clear picture emerges.

The development of all three experience groups from first to fourth grade follows the general trend described above, i.e. the mean number of components strongly increases. However, the degree of improvement differs, especially between mono- and bilingual experience group. The monolingual group has a much lower mean score in first grade than the bilingual group, but its performance also improves much more dramatically (Fig. 5.8 & Tab. 5.5): The monolingual group's fourth grade mean is almost four times higher than its first grade mean (2,5 vs. 9,8). The bilingual group's fourth grade mean, on the other hand, is "only" about 1,5 times higher (6,1 vs. 10). As a result of these unequal increase rates, virtually no difference between monolingual and bilingual preschool experience group can be observed in fourth grade with respect to the mean number of components (9,8 and 10, respectively).

At the same time the interindividual differences as measured by the standard deviation decrease in both experience groups – from a ratio of 45% of the mean in first to 18% in fourth grade for the bili group and from 72% to 16% for the mono group (cf. also Tab. 5.5). Thus, bili fourth graders have a marginally higher variation in scores. This is also reflected by their slightly larger range (6 components as opposed to 5 components in the mono group (cf. Tab. 5.5) and the frequency distribution, which is given in Fig. 5.10.[6]

[6] However, the lowest score of 6 components is realized by only one child and none of the bili children in fourth grade realizes 7 components (cf. Tab. 5.6). Thus, a score of 8 or more components may be more typical for this group.

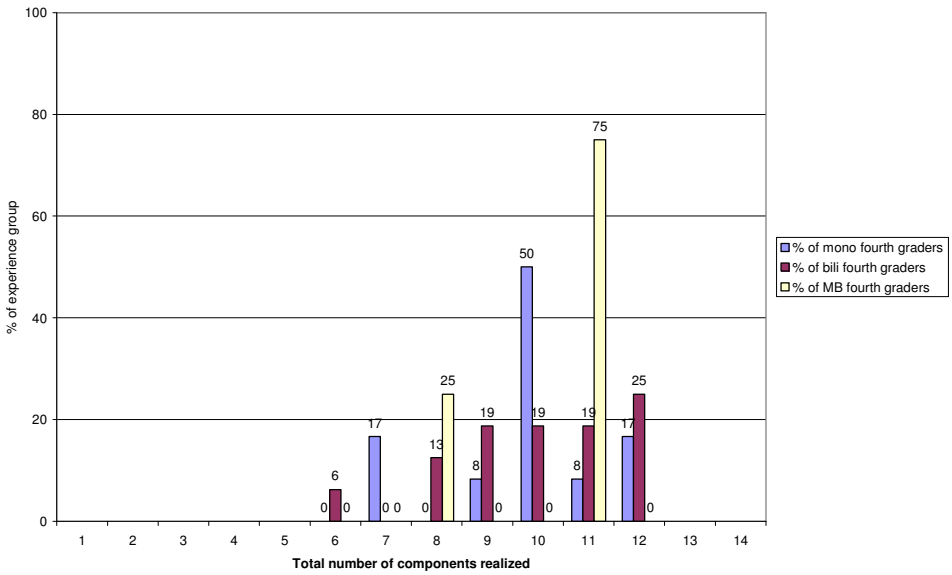

Fig. 5.10 Distribution of the total number of components by experience group in fourth grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children who attended a monolingual group in a preschool offering also bilingual groups.

The frequency distribution shows that the results of bili fourth graders are again not only more varied but also more evenly distributed, while the mono group's results are dominated by a cluster of scores at 10 components. Overall, however, the differences in distribution between mono and bili fourth graders are marginal.

The MB group again seems to take an intermediate position in its development from first to fourth grade (cf. Tab. 5.5 & Fig. 5.8). That is, its fourth grade mean is approximately three times higher than its first grade mean (3,7 vs. 10,3). At the same time the MB participants' interindividual variation also decreases. The relatively large homogeneity in fourth grade results is reflected in the MB fourth graders' distribution (Fig. 5.10), which is split up between 8 (25%) and 11 components (75%). In terms of its fourth grade results, finally, the MB group's performance closely resembles that of the two other groups. That is, all three experience groups perform very similarly in fourth grade.

A final comparison needs to be made between each experience group's first and fourth grade distribution. Thus, the distribution of the total number of narrative components as a function of grade and L2 preschool experience is depicted in Fig. 5.11 for the mono, in Fig. 5.12 for the bili and in Fig. 5.13 for the MB group.

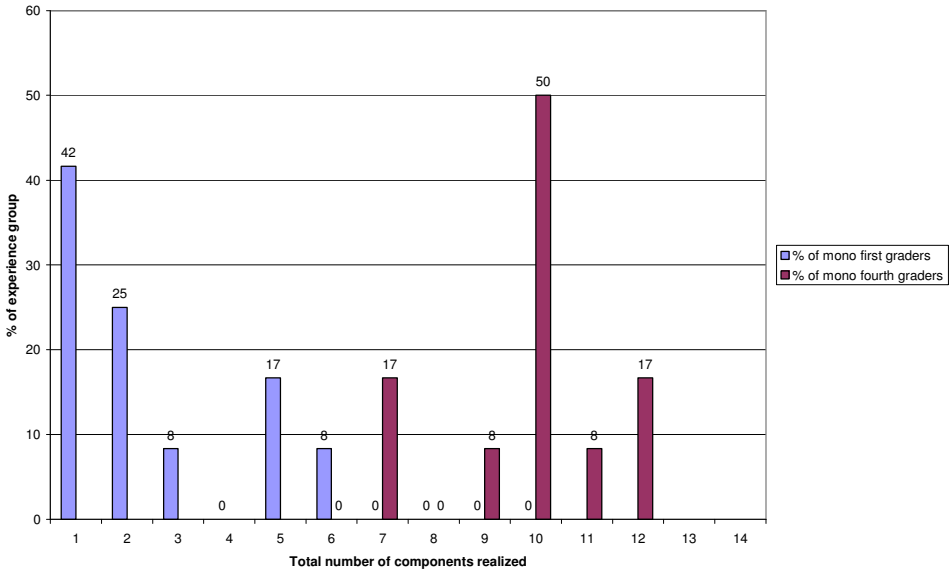

Fig. 5.11 Distribution of the total number of components by grade for participants with exclusively monolingual preschool experience (mono group)

Mono first and fourth grade results are clearly distinguishable: Mono first graders score 6 or fewer components, while mono fourth graders score 7 or more components; there is no overlap in scores.

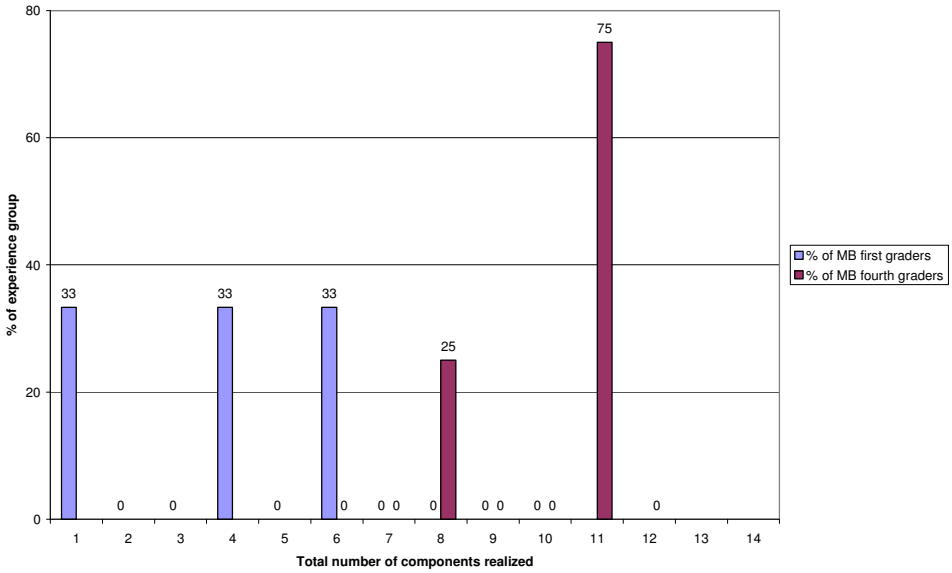

Fig. 5.12 Distribution of the total number of components by grade for children who attended a monolingual group in a preschool offering also bilingual groups (MB group)

The same is true for the MB group's results in first and fourth grade: First graders produce up to 6 components, while fourth graders realize 8 or more components.

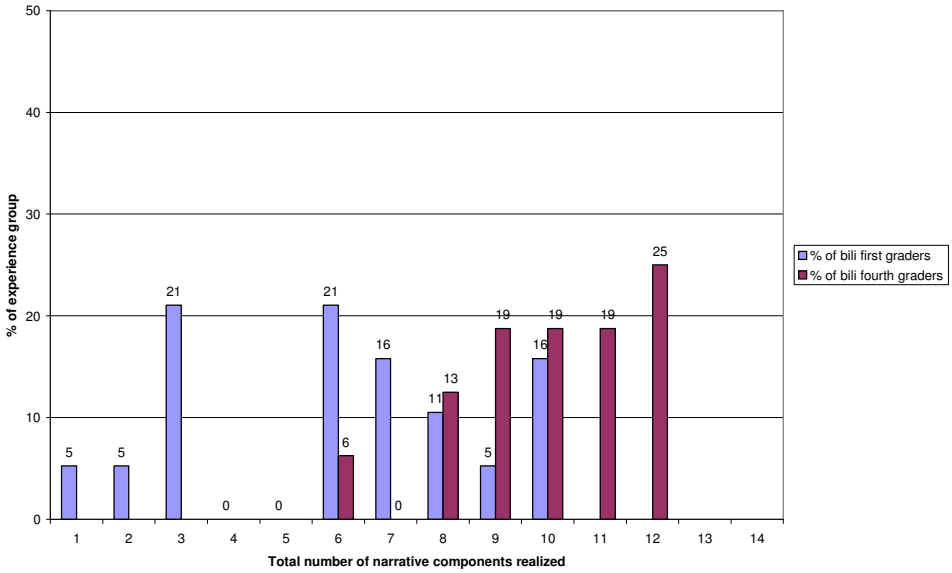

Fig. 5.13 Distribution of the total number of components by grade for children with bilingual preschool experience (bili group)

With respect to bili first and fourth grade scores a different picture emerges. While scores of three components or less are exclusive to bili first and scores of 11 components or more exclusive to bili fourth graders, there is an overlap in results from 6 to 10 components, i.e. from the bili group's fourth grade minimum to the first grade maximum. 68% of first (13 of 19) and 56% of fourth graders (9 of 16) achieved a total number of components within this range. The one bili fourth grader (6%, C5-G4-20) producing only six components could be considered a statistical outlier, however, since all other bili fourth graders realized 8 or more components. So the overlap is actually smaller and ranges from 8 to 10 components with 32% of bili first (6 of 19) and 50% of fourth graders (8 of 16) achieving a total number of components within this range. Nevertheless, a substantial overlap remains, which indicates that many of the bili first graders already perform like fourth graders, while this is not the case for any of the mono or MB first graders.

To sum up, a comparison of the total number of components by experience group found that all three groups under investigation largely followed the overall trends identified earlier in this section. That is, their mean number of components strongly increased from first to fourth grade while the interindividual variation decreased. When comparing participants with and without L2 preschool experience, however, it was found that experience group influences the results in several respects. Thus, bili participants outperformed mono participants in terms of the mean number of components in first grade and in terms of an overlap of scores between

first and fourth grade. At the same time bili first graders' interindividual differences were somewhat lower. The distribution of scores also showed an impact of L2 preschool experience in that a score of 3 components or less was found most typical for mono and a score of 6 or more for bili first graders; this pattern underlines the large difference between experience groups in first grade. However, mono participants' mean score increased far more from first to fourth grade and their interindividual variation decreased more strongly. Since additionally the small MB group was found to take an intermediate position in first grade results as well as in development, virtually no differences attributable to L2 preschool experience were found by the end of fourth grade. This lack of differences also held true for the distribution, which indicated that 7 or more components were most typical for mono and 8 or more for bili participants.

5.1.2 Total number of components: Statistical results

A general linear model analysis of variance was conducted to statistically test main and interaction effects of grade, sex, and L2 experience group on the total number of components realized (cf. ch. 4.2.3). It showed a very highly significant main effect of grade ($F(1, 54)=58,91$, $p<0,001$), which confirms the observed difference between the mean number of realizations in grade one and four described above. This effect of grade accounted for 52% of the overall variance ($\eta_p^2=0,52$). In addition, a significant main effect of experience ($F(2, 54)=3,54$, $p<0,05$) was found, which accounts, however, for only 12% of the overall variance ($\eta_p^2=0,12$). Participants' sex, on the other hand, was not found to be a significant influence factor ($F(1, 54)=1,25$, ns). A significant interaction effect of grade and experience ($F(2, 54)=3,78$, $p<0,05$, $\eta_p^2= 0,12$) was also found, which qualifies the two main effects. That is, it indicates that the effect of experience group was not the same for the two grades. To further explore the difference between mono and bili results, separate t-tests were conducted between mono and bili group for grade 1 and 4. They confirmed that the experience group effect is attributable to the strong difference found for first grade ($t(29)=3,8$, $p<0,05$), where experience accounts for 33% of the variation ($\eta^2=0,33$), whereas by grade four there is no statistically significant difference attributable to L2 preschool experience ($t(26)=0,26$, ns).

Thus, the statistical analysis fully confirmed the observed results showing a very strong effect of grade on variation in the total number of components as well as a

strong effect of L2 experience on first grade results. In fourth grade, however, experience group was not found to be a significant factor anymore. Additionally, the statistical analysis showed that participants' sex does not have any significant influence on the total number of components realized.

5.1.3 Summary: Total number of narrative components

The present section aimed at answering the question how coherent participants' stories were as measured by the number of narrative components realized in first and fourth grade and whether there were any quantitative differences attributable to grade, sex or experience. Two hypotheses had been formulated on the basis of previous studies (cf. ch. 3.6):

(1) Participants' stories become more coherent from first to fourth grade as measured by the number of narrative components

(2) There are qualitative differences in narrative coherence between grades as measured by differences in frequency among the individual narrative components.

First of all, it was found that all first graders were able to produce at least one component and that none of the participants produced all 14 components – even in fourth grade the highest score was 12. Apart from this, two developmental trends were identified in the data, namely a strong increase in the mean number of narrative components from first to fourth grade as well as a strong decrease in interindividual variation. Thus, participants' mean number of components doubled from 4,6 in first to 10 components in fourth grade; the statistical analysis showed that this increase was very highly significant and explained slightly more than half of the variation in results (52%). At the same time the heterogeneity of results decreased to roughly 25% of its first grade value (from 65% to 16% of the mean).

With respect to an influence of sex and experience group the following was found: There were no significant differences between male and female participants in either grade. At the same time both groups' results followed the overall pattern of an increase in mean score and a decrease in heterogeneity of results from first to fourth grade. This was also true for all three experience groups under investigation. However, the comparison between mono and bili group found that L2 preschool experience had a significant impact on first grade results, where the bili group produced a significantly higher mean number of components (6,1 vs. 2,5) and

performed more homogeneously. The statistical analysis showed that L2 preschool experience accounts for 33% of the variation in first grade. Due to the mono groups' stronger increase in the mean number of components and decrease in interindividual variation, however, the differences between monos and bilis had disappeared by the end of fourth grade. The small MB group was found to take an intermediate position in results as well as in development so that in fourth grade there were no longer any differences attributable to experience group.

The distribution of first and fourth grade scores also qualified the overall results, since it showed a large overlap between first and fourth grade scores. That is, 44% of first graders produced results that could equally have been produced by a fourth grader, since these lay above the fourth grade minimum. However, this overlap was again influenced by L2 preschool experience. That is, mono and MB results showed no overlaps at all, while there was a strong overlap between bili first and fourth grade results. Thus, the overlap of first and fourth grade scores was found to be entirely attributable to participants with L2 preschool experience.

With the help of the respective distributions it was also established that a score of six components or less is most typical for first graders and a score of ten or more for fourth graders. Male and female participants' distributions roughly agreed with this pattern, while L2 preschool experience was again found to have a significant impact in first grade: The majority of bili first graders (68%) produced six or more components, while only about one third of them produced three components or fewer. Mono first graders, on the other hand, realized six components or fewer and most typically three or fewer (75%). In fourth grade, both experience groups roughly corresponded to the overall pattern and were very similar to each other.

To sum up, the mean coherence of participants' stories – as measured by the total number of components – increased from first to fourth grade and at the same time the interindividual differences in coherence became much smaller. However, the increase in coherence was qualified by a strong effect of experience group in that most first graders with bilingual preschool experience produced stories in the range of fourth graders. Children without such prior L2 experience, on the other hand, caught up until the end of fourth grade through a comparatively stronger increase in coherence.

5.2 Individual narrative components

This section aims at answering the question whether there are any qualitative differences in coherence as measured by the frequency of the 14 individual components under investigation and whether these differences are attributable to grade, sex or L2 preschool experience.

5.2.1 Individual narrative components: Observed and statistical results

Fig. 5.14 shows all 14 narrative components and their relative frequencies in first and fourth grade in the order of occurrence in the story; Tab. 5.7 gives the corresponding values.

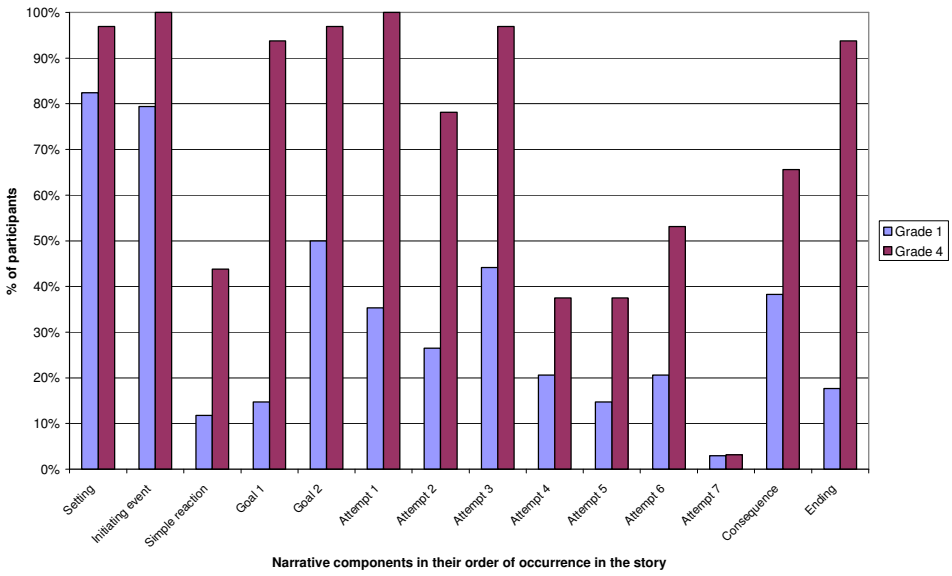

Fig. 5.14 Frequencies of the 14 individual narrative components by grade

	Grade 1 (N=34)		Grade 4 (N=32)	
	N	%	N	%
Setting	28	82	31	97
Initiating event	27	79	32	100
Simple reaction	4	12	14	44
Goal 1	5	15	30	94
Goal 2	17	50	31	97
Attempt 1	12	35	32	100
Attempt 2	9	26	25	78
Attempt 3	15	44	31	97
Attempt 4	7	21	12	38
Attempt 5	5	15	12	38
Attempt 6	7	21	17	53
Attempt 7	1	3	1	3
Consequence	13	38	21	66
Ending	6	18	30	94

Tab. 5.7 Frequencies of the 14 individual narrative components by grade

All of the narrative components are realized by at least one of the children as early as first grade. At the same time only SETTING (82%) and INITIATING EVENT (79%) are produced by the majority of first graders. All other 12 components are observed in 50% or less. GOAL 2 ("find the frog") is at least realized by half of the first graders (50%) – albeit only implicitly[7] – while the remaining components are produced by only up to 44% of them. Between 20% and 44% of first graders realize (in descending order) ATTEMPT 3 (44%), CONSEQUENCE (38%), ATTEMPTs 1 (35%), 2 (26%), 4 and 6 (both 21%). Less than 20% of first graders were able to produce ENDING (18%), GOAL 1 ("recover/ replace the frog", 15%), ATTEMPT 5 (15%) and SIMPLE REACTION (12%). ATTEMPT 7 ("lake attempt", 3%) is realized least often, namely by only one child (C8-G1-21).

As stated in ch. 3, some narrative components seem to be easier to acquire than others in the sense that they are produced from an earlier age on. Thus, the following order of difficulty emerges in first grade as measured by frequency of realizations (Fig. 5.15):

[7] I.e. as deducible from the realization of INITIATING EVENT and two ATTEMPTs.

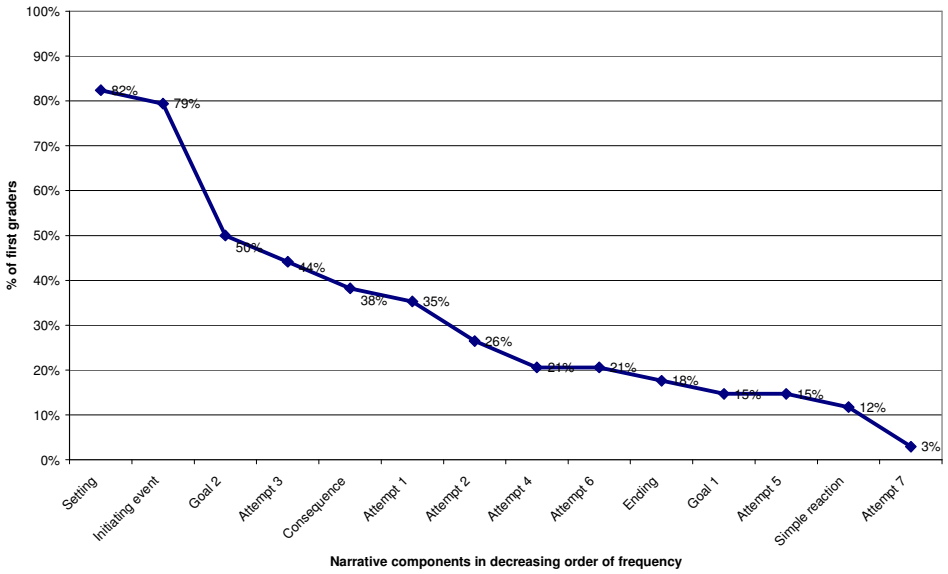

Fig. 5.15 Frequency of each narrative component in first grade in descending order

To sum up, most first graders describe (a) who they are talking about, i.e. the protagonists, and (b) the event that sets up the "problem" that needs to be solved in the ensuing development sequence of the story (cf. ch. 2.3), i.e. the escape of the frog. Even though they use few other narrative components, however, there is evidence for an emerging problem-resolution structure besides the realization of the initiating event: The narrative components following SETTING and INITIATING EVENT in frequency are GOAL 2, which is inferable in half of the stories, ATTEMPTs 3 and 1, which correspond to the story's two possibilities for encoding the beginning of the GOAL PATH (i.e. the recursive ATTEMPT sequence), and CONSEQUENCE, which is one of the two possibilities to end the GOAL PATH. Thus, first graders seem to be on the right track, even though the latter narrative components are encoded by a very low percentage of them.

The frequency of all narrative components besides ATTEMPT 7 increases from first to fourth grade albeit to very different degrees, which are somewhat conditioned by components' first grade frequency – SETTING, for example, is produced by 82% of first graders, which does not leave much room for an increase. The great majority of these observed increase rates were found either to be statistically significant or at least indicating a trend towards significance.

Fig. 5.16 gives the individual components' increase rates, i.e. the difference between their first and fourth grade frequency.

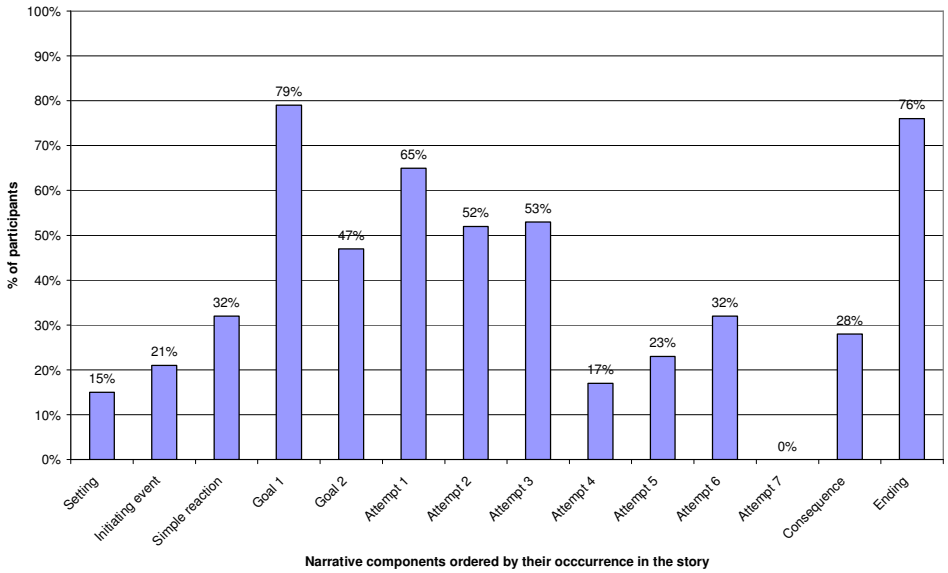

Fig. 5.16 Increase in frequency from first to fourth grade for each narrative component

Fig. 5.16 shows the following order in increase rates: The strongest increase rates observed, which were all statistically very highly significant, were found for GOAL 1 (79%, $\chi^2(1)=34,7$, Fisher's $p<0,001$, $\Phi=0,77$), ENDING (76%, $\chi^2(1)=34,7$, Fisher's $p<0,001$, $\Phi=0,77$), ATTEMPT 1 (65%, $\chi^2(1)=25,3$, Fisher's $p<0,001$, $\Phi=0,66$), ATTEMPT 3 (53%, $\chi^2(1)=18,2$, Fisher's $p<0,001$, $\Phi=0,56$), ATTEMPT 2 (52%, $\chi^2(1)=14,5$, Fisher's $p<0,001$, $\Phi=0,5$) and GOAL 2 (47%, $\chi^2(1)=14,95$, Fisher's $p<0,001$, $\Phi=0,5$). Additionally, the respective effect sizes show the strong influence of grade since they indicate that grade explains no less than 50% and as much as 77% of the variation in results for these components. Even most of the components with lower increase rates, however, differed significantly between first and fourth grade. Thus, first and fourth grade results of ATTEMPT 6 (32%, $\chi^2(1)=6,2$, Fisher's $p<0,05$, $\Phi=0,32$), SIMPLE REACTION (32%, $\chi^2(1)=6,7$, Fisher's $p<0,05$, $\Phi=0,34$) and INITIATING EVENT (21%, $\chi^2(1)=6,03$, Fisher's $p<0,05$, $\Phi=0,32$) were statistically significant, even if grade explained a comparatively lower percentage of these components' variation, namely 32% and 34%. A trend toward significance was found for CONSEQUENCE (28%, $\chi^2(1)=3,9$, Fisher's $p=0,07$, ns) and ATTEMPT 4 (17%, $\chi^2(1)=3,8$, Fisher's $p=0,09$, ns). Only the observed increase in ATTEMPT 5 (23%, $\chi^2(1)=2,97$, Fisher's $p=0,13$, ns) and SETTING (15%, $\chi^2(1)=3,5$, Fisher's $p=0,11$, ns) was not statistically significant.

In fourth grade, large differences in frequency remain between the individual components (Fig. 5.14 & Tab. 5.7) and the following order of difficulty of realization emerges as measured by the frequency of realization (Fig. 5.17):

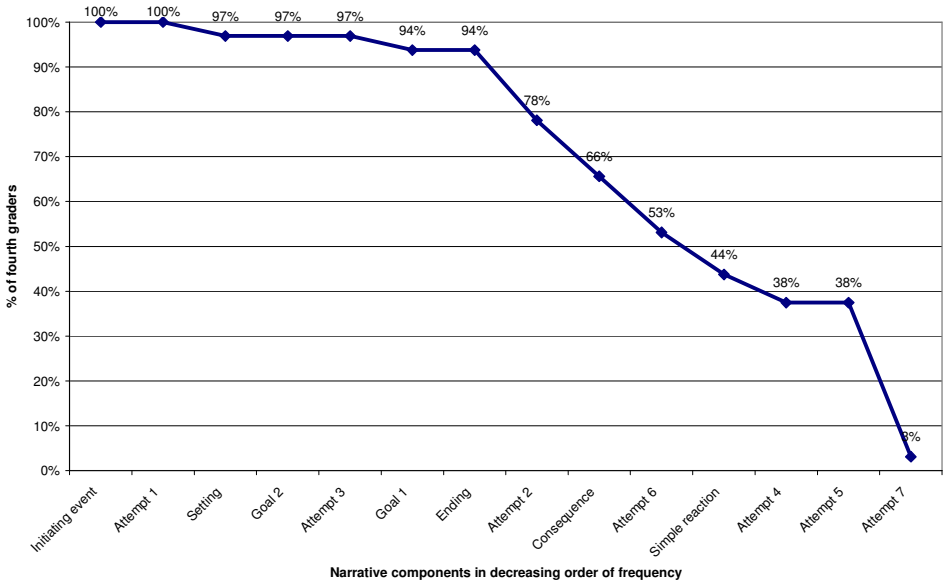

Fig. 5.17 Frequency of each narrative component in fourth grade in descending order

ATTEMPT 1 and INITIATING EVENT are realized by all fourth graders (100%) and a further five narrative components are realized by over 90% of them, namely SETTING (97%), GOAL 2 (97%), ATTEMPT 3 (97%), GOAL 1 (94%) and ENDING (94%).[8] In absolute numbers, 97% correspond to merely one and 94% to two of the fourth graders not realizing the respective component. That is, half of the fourteen narrative components are realized by almost all of the fourth graders. The remaining seven components do not show any similarities in frequency. 78% of the fourth graders realize the "window attempt", i.e. ATTEMPT 2. Still over half of the fourth graders realize CONSEQUENCE (66%) and ATTEMPT 6 (53%). Not even half of them, however, realize SIMPLE REACTION (44%) and ATTEMPTs 4 and 5 (both 38%). Narrative component ATTEMPT 7 is again realized by only one participant (3%, C5-G4-10). Consequently, the following order of difficulty arises in fourth as compared to first grade:

[8] The high frequency of goals 1 and 2 is attributable to their implicit realization as inferable from these components; only GOAL 2 is also realized explicitly and then by a mere 11 of the 32 fourth graders, i.e. 34% of them.

Order	Grade 1	Grade 4
1.	Setting	Initiating event
2.	Initiating event	Attempt 1
3.	Goal 2	Setting
4.	Attempt 3	Goal 2
5.	Consequence	Attempt 3
6.	Attempt 1	Goal 1
7.	Attempt 2	Ending
8.	Attempt 4	Attempt 2
9.	Attempt 6	Consequence
10.	Ending	Attempt 6
11.	Goal 1	Simple reaction
12.	Attempt 5	Attempt 4
13.	Simple reaction	Attempt 5
14.	Attempt 7	Attempt 7

Tab. 5.8 Order of difficulty for first and fourth graders. Note: Frequencies above 90% are shaded in green, frequencies of 75% or above in yellow and those of 50% or above in blue.

Tab. 5.8 shows that the order of difficulty in first and fourth grade is not very different, if a certain interchangeability in ATTEMPTs 2 to 7 as well as CONSEQUENCE and ENDING is allowed for. The main difference then seems to be a quantitative one: Almost all fourth graders (over 90%) describe not only – as most first graders do – (a) protagonists (SETTING) and (b) the problem that triggers the attempt sequence (INITIATING EVENT). Besides these, they add further components of a global narrative structure in the sense of the narrative schema outlined in ch. 2. Thus, the great majority of fourth graders (over 90%) additionally realize (c) the initiation of search theme and repeated attempt sequence (ATTEMPTs 1 & 3) and (d) the completion of the search theme (ENDING). Additionally, the underlying GOALs motivating the story characters can be inferred in almost all of the stories, even if they are still rarely realized explicitly.[9]

It may seem that even fourth graders produced only half of all the components necessary for a global narrative structure. However, the remaining seven components can be considered optional and thus dispensable: No CONSEQUENCE is necessary if the ENDING – which is additionally more salient – is realized. Similarly, ATTEMPTs 1 and 3 initiate the search inside and outside, respectively, and participants may see the search theme as sufficiently established after mentioning these two components (cf. Bamberg & Marchman 1994). Therefore, any additional mentions of the search would rather become a matter of personal style and choice to elaborate than of necessity for a global story structure and theme. Similarly, SIMPLE REACTION is a category that is largely optional and usually left implicit even in adult stories (cf. ch. 2). Thus, the great majority of fourth graders was found to produce the

[9] GOAL 1 is never realized explicitly and GOAL 2 only 11 times, i.e. by 34% of fourth graders.

components most important for a globally structured narrative, while leaving out components which fulfill more of an elaborative function.

To sum up, all 14 narrative components were already produced in first grade but almost all components' frequency increased from first to fourth grade. Thus, only two components were realized by the majority of first graders (over 75%) but seven by almost all fourth graders (over 90%). The statistical analysis showed that the increase in frequency was significant for most of the components. With the help of the respective frequencies an order of difficulty of realization was established, which showed that the development from first to fourth grade went hand in hand with a development towards a global narrative structure. That is, the beginnings of such a structure were already discernable in first grade but only in fourth grade the necessary components were produced by the great majority of participants.

5.2.1.1 Individual narrative components by sex

The realization of the individual components by male and female participants in first and fourth grade is presented in Fig. 5.1 and the corresponding values in Tab. 5.9. Some observable differences exist, which will be described in the following. However, none of these observed differences were statistically significant.

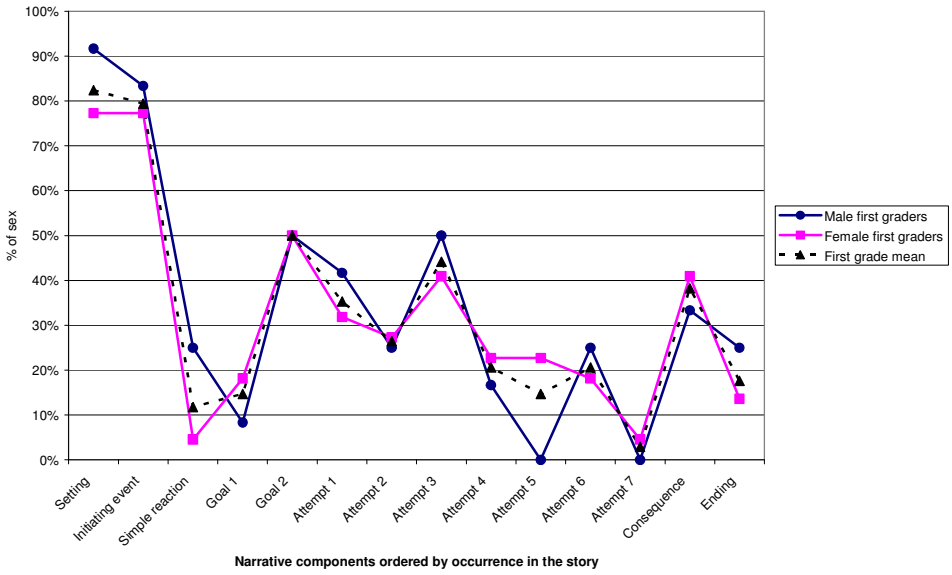

Fig. 5.18 Frequency of each narrative component by sex in first grade

	Grade 1				Grade 4			
	Male (N=12)		Female (N=22)		Male (N=10)		Female (N=22)	
	N	%	N	%	N	%	N	%
Setting	11	92	17	77	10	100	21	95
Initiating event	10	83	17	77	10	100	22	100
Simple reaction	3	25	1	5	5	50	9	41
Goal 1	1	8	4	18	10	100	20	91
Goal 2	6	50	11	50	10	100	21	95
Attempt 1	5	42	7	32	10	100	22	100
Attempt 2	3	25	6	27	8	80	17	77
Attempt 3	6	50	9	41	10	100	21	95
Attempt 4	2	17	5	23	4	40	8	36
Attempt 5	0	0	5	23	2	20	10	45
Attempt 6	3	25	4	18	6	60	11	50
Attempt 7	0	0	1	5	1	10	0	0
Consequence	4	33	9	41	7	70	14	64
Ending	3	25	3	14	10	100	20	91

Tab. 5.9 Frequency of each narrative component by sex and grade

Neither sex performs very differently from the mean in the realization of the individual narrative components in first grade. However, female first graders produced all of the narrative components, while none of the males realized ATTEMPTs 5 and 7. Clearly more[10] male than female first graders, on the other hand, realized SETTING (92% vs. 77%), SIMPLE REACTION (25% vs. 5%), ATTEMPT 1 (42% vs. 32%), and ENDING (24% vs. 14%).[11] Slightly more[12] male participants produced ATTEMPT 3 (50% vs. 41%), ATTEMPT 6 (25% vs. 18%) and INITIATING EVENT (83% vs. 77%).[13] Female first graders, on the other hand, realized ATTEMPT 5 (23% vs. 0%) as well as GOAL 1 (18% vs. 8%) clearly more often and ATTEMPT 4 (23% vs. 17%) as well as CONSEQUENCE (41% vs. 33%) slightly more often.[14] Male first graders thus seem

[10] The expression 'clearly more' is used in the following to indicate that observed results differ by 10% or more. This does not necessarily mean that the difference is also statistically significant.

[11] The statistical results are: SETTING ($\chi^2(1)=0,83$, ns), SIMPLE REACTION ($\chi^2(1)=3,84$, Fisher's $p=0,09$), which could be interpreted as a trend towards significance, ATTEMPT 1 ($\chi^2(1)=0,79$, ns), and ENDING ($\chi^2(1)=0,16$, ns).

[12] A difference of less than 10%.

[13] On all of the components mentioned they simultaneously scored higher than the mean. The statistical results are: ATTEMPT 3 ($\chi^2(1)=1,31$, ns), ATTEMPT 6 ($\chi^2(1)=1,07$, ns) and INITIATING EVENT ($\chi^2(1)=0,83$, ns).

[14] The statistical results are: GOAL1 ($\chi^2(1)=0,41$, ns), ATTEMPT 4 ($\chi^2(1)=0,004$, ns), CONSEQUENCE ($\chi^2(1)=0,47$, ns). ATTEMPT 5 could not be tested reliably due to expected cell frequencies below 1 (cf. ch. 4.2.3).

to have a slight advantage on some of the components important for a global narrative structure (SETTING, INITIATING EVENT, ATTEMPT 1 and/ or 3).[15]

However, the difficulty of realization in first grade, which is shown in Tab. 5.10, is quite similar for both sexes.

Order	Male first graders	Female first graders
1.	Setting	Setting
2.	Initiating event	Initiating event
3.	Goal 2	Goal 2
4.	Attempt 3	Attempt 3
5.	Attempt 1	Consequence
6.	Consequence	Attempt 1
7.	Simple reaction	Attempt 2
8.	Attempt 2	Attempt 4
9.	Attempt 6	Attempt 5
10.	Ending	Goal 1
11.	Attempt 4	Attempt 6
12.	Goal 1	Ending
13.	Attempt 5	Simple reaction
14.	(Attempt 7)	Attempt 7

Tab. 5.10 Order of difficulty for first graders by sex. Note: Frequencies above 90% are shaded in green, frequencies of 75% or above in yellow and those of 50% or above in blue; parentheses indicate that components were not produced by any of the group members.

The majority of male and female first graders realize SETTING and INITIATING EVENT and half of them realize GOAL 2, while all other components are realized by less than half of males and females; in this they follow the overall results described above. The remaining components are realized in a similar order of difficulty, which also corresponds closely to the overall order of difficulty described earlier. The only notable exception – if one assumes a certain degree of interchangeability in the ATTEMPTs – is SIMPLE REACTION (25% vs. 5%, i.e. 7th rank vs. 13th rank).

Both male and female participants increase their performance from first to fourth grade for virtually all narrative components. The only exception is ATTEMPT 7, which is produced by a female participant in first and by a male participant in fourth grade. The increase rates of both sexes roughly follow the overall increase rates discussed earlier, i.e. they vary strongly from component to component. In fourth grade female and male participants thus perform again very similar to the mean; this is evident from Fig. 5.19, which shows the frequency of each narrative component by sex in fourth grade.

[15] As explained in ch. 2, either first or third attempt can serve to introduce the search theme. The ENDING component was not included since it can be compensated by CONSEQUENCE, which is realized slightly more often by female first graders.

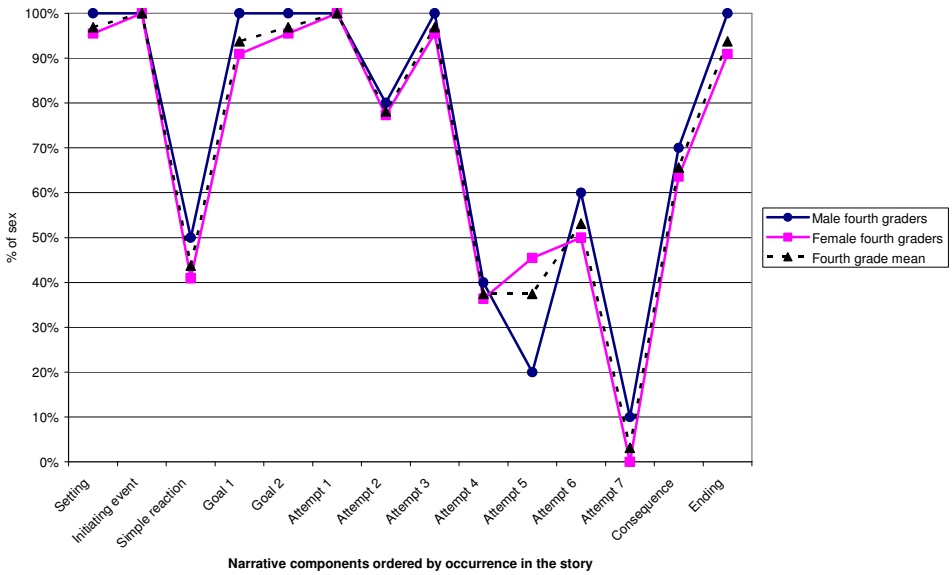

Fig. 5.19 Frequency of each narrative component by sex in fourth grade

Female participants' performance resembles the mean in all but ATTEMPT 5, while male fourth graders realize several components not only slightly more often than the mean but also slightly more often than female fourth graders. That is, male fourth graders slightly more often realized SETTING (100% vs. 95%), SIMPLE REACTION (50% vs. 41%), GOAL 1 (100% vs. 91%), GOAL 2 (100% vs. 95%), ATTEMPT 6 (60% vs. 50%), ATTEMPT 7 (10% vs. 0%), CONSEQUENCE (70% vs. 64%) and ENDING (100% vs. 91%).[16] In absolute numbers these differences translate into only one or two participants (not) realizing the corresponding component (Tab. 5.9) and thus they seem negligible. Female fourth graders realized one component, ATTEMPT 5, clearly more often than their male counterparts (45% vs. 20%)[17] but performed either the same or slightly lower for all other components. At the same time the order of difficulty as measured by the frequency of realization is again very similar for both sexes (Tab. 5.11):

[16] The statistical results are: SIMPLE REACTION ($\chi^2(1)=0,78$, ns), ATTEMPT 6 (60% vs. 50%, $\chi^2(1)=0,19$, ns) and CONSEQUENCE (70% vs. 64%, $\chi^2(1)=0,21$, ns). SETTING, GOAL 1 and 2, ATTEMPT 7 and ENDING could not be tested reliably due to expected cell frequencies below 1 (cf. ch. 4.2.3).

[17] ATTEMPT 5 ($\chi^2(1)=1,87$, ns).

Order	Male fourth graders	Female fourth graders
1.	Setting	Initiating event
2.	Initiating event	Attempt 1
3.	Goal 1	Goal 2
4.	Goal 2	Attempt 3
5.	Attempt 1	Setting
6.	Attempt 3	Goal 1
7.	Ending	Ending
8.	Attempt 2	Attempt 2
9.	Consequence	Consequence
10.	Attempt 6	Attempt 6
11.	Simple reaction	Attempt 5
12.	Attempt 4	Simple reaction
13.	(Attempt 5)	Attempt 4
14.	(Attempt 7)	Attempt 7

Tab. 5.11 Order of difficulty for fourth graders by sex. Note: Frequencies above 90% are shaded in green, frequencies of 75% or above in yellow and those of 50% or above in blue; parentheses indicate that components were not produced by any of the group members

To sum up, there were almost no differences between male and female participants in either grade with respect to the realization of the individual narrative components – even more so when keeping in mind a possible group size effect. Instead both groups closely follow the overall pattern. The statistical analysis also supported the conclusion that the differences between male and female participants are negligible, since none of the differences tested were significant.

5.2.1.2 Individual narrative components by experience group

The realization of each narrative component by experience group in first grade is shown in Fig. 5.20; the corresponding values are given in Tab. 5.12. A main comparison will again be made between participants with and without L2 preschool experience (bili vs. mono), while the MB group will be regarded separately due to its small size and thus limited representativeness.

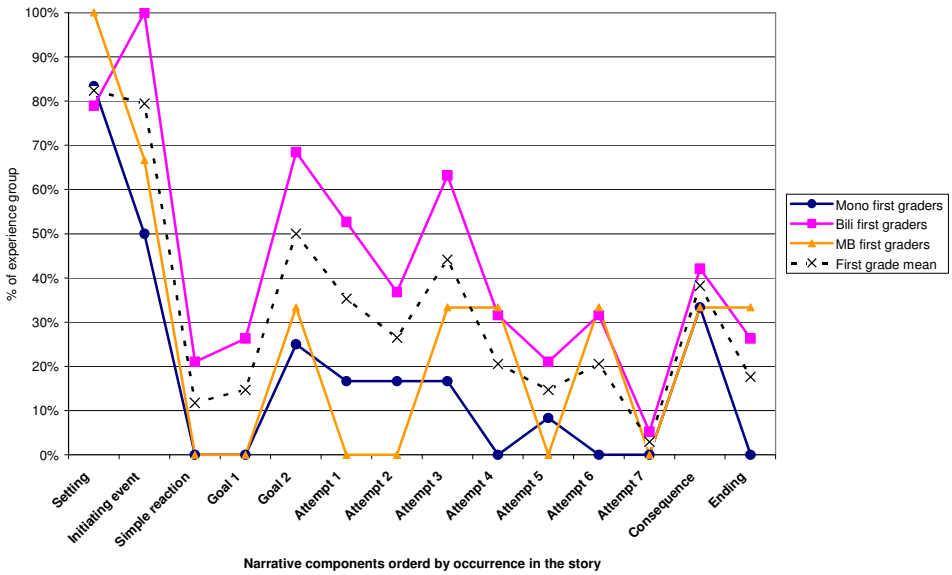

Fig. 5.20 Frequency of each narrative component in first grade by experience group. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children who attended a monolingual group in a preschool offering also bilingual groups.

	Grade 1						Grade 4					
	Mono (N=12)		Bili (N=19)		MB (N=3)		Mono (N=12)		Bili (N=16)		MB (N=4)	
	N	%	N	%	N	%	N	%	N	%	N	%
Setting	10	83	15	79	3	100	12	100	15	94	4	100
Initiating event	6	50	19	100	2	67	12	100	16	100	4	100
Simple reaction	0	0	4	21	0	00	4	33	8	50	2	50
Goal 1	0	0	5	26	0	00	10	83	16	100	4	100
Goal 2	3	25	13	68	1	33	12	100	15	94	4	100
Attempt 1	2	17	10	53	0	0	12	100	16	100	4	100
Attempt 2	2	17	7	37	0	0	9	75	13	81	3	75
Attempt 3	2	17	12	63	1	33	12	100	15	94	4	100
Attempt 4	0	0	6	32	1	33	6	50	6	38	0	0
Attempt 5	1	8	4	21	0	00	7	58	3	19	2	50
Attempt 6	0	0	6	32	1	33	4	33	10	63	3	75
Attempt 7	0	0	1	5	0	00	1	8	0	0	0	0
Consequence	4	33	8	42	1	33	7	58	11	69	3	75
Ending	0	0	5	26	1	33	10	83	16	100	4	100

Tab. 5.12 Frequency of each narrative component in first and fourth grade by experience group. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children who attended a monolingual group in a preschool offering also bilingual groups.

All narrative components are produced by at least some of the bili first graders, while none of the mono first graders realized SIMPLE REACTION, GOAL 1, ATTEMPTs 4,

6 and 7 and the ENDING. Additionally, Fig. 5.20 and Tab. 5.12 show that the higher mean number of components achieved by first graders with bilingual preschool experience (cf. ch. 5.1.1.3) is due to a consistently higher number of realizations across all narrative components. Thus, the bili group outperforms the mono group on all but the SETTING, which is realized slightly more often by the monos (mono 83% vs. bili 79%) – statistically this difference was also not significant ($\chi^2(1)=0,76$, Fisher's $p=1,0$, ns). Little observable difference between the two experience groups was found for ATTEMPT 7 (bili 5% vs. mono 0%) and CONSEQUENCE (bili 42% vs. mono 33%). The difference in ATTEMPT 7 corresponds to one bilingual first grader realizing this component, however, which cannot be considered representative,[18] while the difference in CONSEQUENCE ($\chi^2(1)=0,63$, Fisher's $p=0,72$, ns) was not statistically significant.

All components besides SETTING, ATTEMPT 7 and CONSEQUENCE are thus realized clearly more often by the bili group, i.e. by at least 10% more bili than mono participants: Over 30% more bilis realize INITIATING EVENT (100% of bilis vs. 50% of monos), GOAL 2 (68% vs. 25%), ATTEMPT 1 (53% vs. 17%), ATTEMPT 3 (63% vs. 17%), ATTEMPT 4 (32% vs. 0%) and ATTEMPT 6 (32% vs. 0%). Over 20% more bilis than monos produce SIMPLE REACTION (21% vs. 0%), GOAL 1 (26% vs. 0%), ATTEMPT 2 (37% vs. 17%) and ENDING (26% vs. 0%). Over 10% more of the bilis realized ATTEMPT 5 (21% vs. 8%). Statistically, the difference between mono and bili group was testable for only five of these components, mostly because the remaining six components were not produced by any of the mono first graders.[19] The difference in realizations for GOAL 2 (68% vs. 25%; $\chi^2(1)=5,55$, Fisher's $p<0,05$, $\Phi=0,42$), and ATTEMPT 3 (63% vs. 17%; $\chi^2(1)=6,42$, Fisher's $p<0,05$, $\Phi=0,46$) was found to be statistically significant. As the respective effect sizes indicate, the experience group effect explains between 42% and 62% of the variation in results for these components, which confirms the strong influence of experience group in first grade. The statistical result for ATTEMPT 1 indicated a trend towards significance (53% vs. 17%; $\chi^2(1)=4$, Fisher's $p=0,065$, ns), while the results for ATTEMPT 2 (37% vs. 17%; $\chi^2(1)=1,45$, Fisher's $p=0,42$, ns) and ATTEMPT 5 (21% vs. 8%; $\chi^2(1)=0,88$, Fisher's $p=0,62$, ns) were non-significant.

[18] The difference in realization of ATTEMPT 7 was not statistically testable due to an expected cell frequency of zero for the mono group.
[19] INITATING EVENT, SIMPLE REACTION, GOAL 1, ATTEMPTs 4 and 6 as well as ENDING could not be tested due to expected cell frequencies below 1.

With respect to the order of difficulty in first grade for each of the main experience groups the following picture emerges:

Order	Mono first graders	Bili first graders
1.	Setting	Initiating event
2.	Initiating event	Setting
3.	Consequence	Goal 2
4.	Goal 2	Attempt 3
5.	Attempt 1	Attempt 1
6.	Attempt 2	Consequence
7.	Attempt 3	Attempt 2
8.	Attempt 5	Attempt 4
9.	(Simple reaction)	Attempt 6
10.	(Goal 1)	Ending
11.	(Attempt 4)	Goal 1
12.	(Attempt 6)	Simple reaction
13.	(Attempt 7)	Attempt 5
14.	(Ending)	Attempt 7

Tab. 5.13 Order of difficulty for first graders by experience group. Note: Frequencies above 90% are shaded in green, frequencies of 75% or above in yellow and those of 50% or above in blue. Parentheses indicate that components were not produced by any of the group members.

Four components are realized by the majority of bili first graders, namely SETTING (79%), INITIATING EVENT (100%), GOAL 2 (68%), and ATTEMPT 3 (63%). ATTEMPT 1 (53%) was realized by a little more than half of bili first graders. In the mono group, the SETTING (83%) was also realized by the majority of participants and the INITIATING EVENT by at least half of them (50%). All other components were produced by 25% or less of mono first graders. Thus, bili first graders show the beginnings of a global narrative structure: The majority of first graders with bilingual preschool experience describes the protagonists, the problem triggering the global problem-resolution structure, and at least one of the two attempts which can be considered the beginning of a global search theme (ATTEMPT 1 and 3)[20]. In the case of the mono group, there is no comparably conclusive evidence for an emerging global narrative structure. Additionally, first graders with bilingual preschool experience tell more elaborate, more detailed stories, i.e. they incorporate not only more (cf. ch. 5.2.1.2) but also a greater variety of narrative components.

MB first graders only coincide in realizing the SETTING (100%); additionally, two of its three members realize the INITIATING EVENT (67%). Otherwise no coherent pattern emerges: C1-G1-11 realizes only the SETTING, while C1-G1-18 realizes also INITIATING EVENT, CONSEQUENCE and ENDING. C1-G1-5, on the other hand, realizes SETTING, INITIATING EVENT and GOAL 2 as well as

[20] These two attempts are realized clearly more often than any of the other attempts, which could either be explained through greater saliency in the elicitation material or – more likely – to being considered more important for the story.

ATTEMPTS 3, 4, and 5. Thus, the MB group performs very heterogeneously not only in the overall number of components but also in the realization of the individual narrative components.

All three experience groups follow the general trend of a higher number of realizations in fourth grade for most of the narrative components. Exceptions are mostly attributable to realizations of 100% in first grade (INITIATING EVENT, SETTING), which leave no room for an increase, or to small changes of one participant more or less in first or fourth grade (ATTEMPT 7, ATTEMPT 4). The fourth grade results by experience group are depicted in Fig. 5.21; for the corresponding values cf. Tab. 5.12 above.

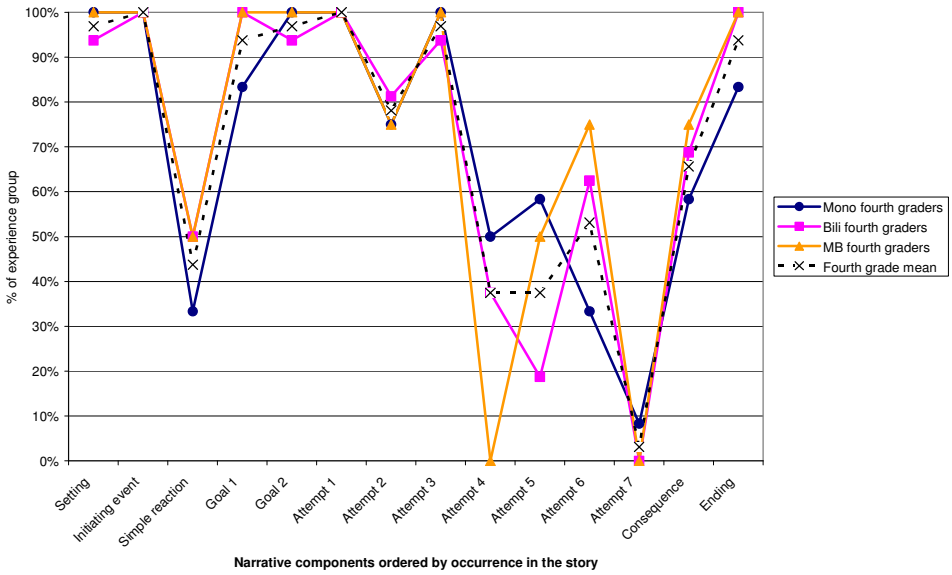

Fig. 5.21 Frequency of each narrative component by experience group in fourth grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

The results of both groups closely resemble each other in fourth grade, i.e. the differences have become quite small compared to first grade, even if a certain advantage of the bili group (over 10% more) can be observed for five of the 14 components, namely SIMPLE REACTION (bili 50% vs. mono 33%), GOAL 1 (100% vs. 83%), ATTEMPT 6 (63% vs. 33%), CONSEQUENCE (69% vs. 58%) and ENDING (100% vs. 83%). A slightly higher performance of the mono group can be observed for ATTEMPT 4 (mono 50% vs. bili 38%) and ATTEMPT 5 (58% vs. 19%). The largest differences are thus found for ATTEMPT 5 (monos 39% more) and ATTEMPT 6 (bilis 30% more). For all other seven components, there are either no observable differences or these differences are below 10% and thus negligible. The

statistical analysis confirms this similarity in results since none of the differences tested were statistically significant.[21]

The order of difficulty (as measured by frequency of realization) for both mono and bili fourth graders is given in Tab. 5.14:

Order	Mono fourth graders	Bili fourth graders
1.	Setting	Initiating event
2.	Initiating event	Goal 1
3.	Goal 2	Attempt 1
4.	Attempt 1	Ending
5.	Attempt 3	Setting
6.	Goal 1	Goal 2
7.	Ending	Attempt 3
8.	Attempt 2	Attempt 2
9.	Consequence	Consequence
10.	Attempt 5	Attempt 6
11.	Attempt 4	Simple reaction
12.	Simple reaction	Attempt 4
13.	Attempt 6	Attempt 5
14.	Attempt 7	Attempt 7

Tab. 5.14 Order of difficulty for fourth graders by experience group. Note: Frequencies above 90% are shaded in green, frequencies of 75% or above in yellow and those of 50% or above in blue; parentheses indicate that components were not produced by any of the group members.

The order of difficulty shows some minor differences between fourth graders with and without bilingual preschool experience in regard to the order of difficulty and the frequency of realization. These differences seem negligible, however, since the same components are included in the seven most frequent realizations and the differences in frequency translate to an absolute number of one or two participants. Thus, a frequency of 94% (SETTING, GOAL 2, ATTEMPT 3) in the bili group (N=16) means that one bili fourth grader did not realize the respective component and a frequency of 83% (GOAL 1, ENDING) in the mono group (N=12) corresponds to two mono fourth graders not realizing the component in question. Furthermore, no statistically significant differences were found between the two experience groups in fourth grade.

MB fourth graders largely perform like the two other experience groups. The only significant exceptions are ATTEMPTs 4, and 7, which are realized by none of the MB fourth graders, ATTEMPT 5, where the MB group (50%) scores higher than the bili (19%) but lower than the mono group (58%), and ATTEMPT 6, where the MB group (75%) outperforms monos (33%) and bilis (63%). Overall, however, these

[21] SIMPLE REACTION ($\chi^2(1)=0{,}78$, ns), ATTEMPT 2 ($\chi^2(1)=0{,}16$, ns), ATTEMPT 4 ($\chi^2(1)=0{,}44$, ns), ATTEMPT 6 ($\chi^2(1)=2{,}33$, ns), CONSEQUENCE ($\chi^2(1)=0{,}32$, ns). Even the result for ATTEMPT 5 missed statistical significance, albeit only barely ($\chi^2(1)=4{,}68$, Fisher's $p=0{,}05$, ns). However, SETTING, GOAL 1 and 2, ATTEMPT 3 and 7, ENDING could not be tested reliably due to expected cell frequencies below 1, INITIATING EVENT and ATTEMPT 1 because they are a constant.

differences are negligible, especially because the ATTEMPTs are equal in rank and largely interchangeable, i.e. differences in their realization need to be considered a matter of style and personal preference.

To sum up, L2 preschool experience was found to have a significant impact on first grade results, where bili clearly outperformed mono participants. Thus, all narrative components were produced by at least one of the bili first graders, while 6 of 14 components (43%) were produced by none of the mono first graders. Qualitatively, only bili first graders showed the beginnings of a global narrative structure and they told more elaborate stories in terms of the variety of components used. No coherent first grade pattern was found for the small MB group. All three experience groups followed the overall trend of an increase in the frequencies of realization and by the end of fourth grade all three groups' results were quantitatively as well as qualitatively very similar to each other and the overall results. That is, the difference between mono and bili group had become negligible until the end of fourth grade where their frequencies of realization and order of difficulty closely resembled each other.

5.2.2 Summary: Individual narrative components

This section aimed at answering the question whether there are any qualitative differences in coherence – as measured by the frequency of the 14 individual components under investigation – attributable to grade, sex or L2 preschool experience. It was found that all narrative components were realized by at least one participant even in first grade and that there was a significant observable and statistical increase in the frequency of realization from first to fourth grade for almost all components. An order of difficulty of realization based on components' frequency showed that only SETTING (82%) and INITIATING EVENT (79%) were produced by the majority of first graders, while the great majority of fourth graders produced additional components necessary for a global narrative structure, especially ATTEMPTs 1 (100%) and 3 (97%) and the ENDING (94%). Thus, the respective orders of difficulty indicated the beginnings of a global structure in first grade. Only the combination of components realized by the great majority of fourth graders, however, actually reflected a global narrative structure, i.e. only fourth graders were found to produce coherent stories.

Participants' sex did not have any significant effect on frequencies or order of difficulty, i.e. male and female participants performed very similarly to each other (as well as to the overall results) in either grade and they followed the overall pattern of an increase in frequencies from first to fourth grade. L2 preschool experience, on the other hand, was found to have a profound influence. However, this influence was limited to first grade. Thus, all narrative components were produced by at least one bili first grader, while six components were realized by none of the mono first graders. Qualitatively, SETTING (79%), INITIATING EVENT (100%), ATTEMPT 1 (53%) and ATTEMPT 3 (63%) were produced by over 50% of bili first graders. Only SETTING (83%) and INITIATING EVENT (50%), on the other hand, were realized by 50% or more of mono first graders, and no coherent pattern emerged for the small group of MB first graders. That is, the evidence for an emerging global narrative structure in first grade is mainly attributable to first graders with L2 preschool experience. In terms of the development from first to fourth grade all three experience groups followed the overall trend of an increase in frequency and by the end of fourth grade the influence of L2 preschool experience had disappeared, i.e. all three groups performed largely the same with respect to components' frequencies and order of difficulty.

To sum up, a general trend of increase in the frequency of realization was found, which applied to almost all components. At the same time L2 preschool experience significantly influenced first grade results in that many of the components necessary for a coherent story were produced by a considerable number of bili first graders as compared to none or few of the other participants. In fourth grade, most components indispensable for a coherent story were produced by the great majority of participants and there were no significant differences between any of the subgroups.

5.3 Index of global narrative structure

In order to more closely investigate the participants' qualitative development in the sense of a global narrative structure, which seems to be in its initial stages by the end of first grade and largely developed by grade 4 (cf. the previous section), a weighted index was constructed. This index of global narrative structure was defined as the sum of the five narrative components most important for representing the

narrative structure of the frog story (cf. ch. 4.2.1). In the following I will first briefly describe the criteria applied for index construction and then the results obtained.

5.3.1 Index construction and reliability

First of all, the SETTING (mention of all three protagonists) was included as it is one of the integral parts of a story according to story grammar (cf. ch. 2.3.2). Secondly, a main focus in the analysis of global structure lies in whether participants establish and conclude a problem-resolution structure around which the story is organized. Since it marks the beginning of this structure, the INITIATING EVENT (escape of the frog) was also incorporated in the index. Additionally, the instantiation of the search (ATTEMPT 1) was included as a crucial indicator for the global problem-resolution structure. Since participants have several possibilities of encoding the first attempt (cf. ch. 4.2.1), this component was redefined as realization of either ATTEMPT 1 or 2 or 3, which are all legitimate options for encoding the initiation of the search.

A second ATTEMPT component was added to the index as an indicator for a sustained search motif (cf. Akinçi et al. 2001, Kuppersmitt & Berman 2001, Berman & Slobin 1994, Bamberg & Marchman 1990) and it was defined as any one attempt following the first attempt realized. However, this component does not have the same weight for global structuring as the other components, since the search theme becomes the default case once it has been instantiated (cf. Bamberg & Marchman 1990, 1991, 1994). Consequently, the second attempt was assigned simple weight, while all other components were assigned double weight.

Finally, a RESOLUTION component was introduced. As pointed out earlier, considerable overlap can exist between OUTCOME and ENDING in one-episode stories such as the frog story. At the same time, both higher-order goal ‘recover or replace the frog’ and lower-order goal ‘find the frog’ (cf. ch. 4.2.1) can serve as the global organizing theme of the frog story. Consequently, either CONSEQUENCE (finding the frog) or ENDING (recovering/ replacing the frog) can conclude the story theme. Thus, outcome and ending were combined into the RESOLUTION (cf. Berman & Slobin 1994), which was defined as the realization of CONSEQUENCE, ENDING or both.

The narrative components composing the global narrative index and their assigned weight are summarized in Tab. 5.15. The index was found to be reliable (Cronbach’s $\alpha = 0{,}72$) and thus the global narrative structure of participants’ texts was

measured by their score on a range from 1 to 9 with nine points signalling a fully developed global narrative structure. Additionally, the index was found to correlate very strongly with the overall number of components realized (r=0,904) with an effect size of r^2=0,82, which means that this "reduced" version of story grammar can be used as a very representative measure also for the total number of components.

Index component	Story grammar component	Weight assigned
Setting	SETTING	double
Initiating event	INITIATING EVENT	double
First attempt	ATTEMPT 1, 2 or 3	double
Subsequent attempt	Any following attempt	single
Resolution	CONSEQUENCE and/or ENDING	double

Tab. 5.15 Index construction: Components and their assigned weight

5.3.2 Global narrative index results

This section complements the results presented for total number of components and individual narrative components by focusing on a combination of components representative for a coherent story. That is, it is investigated whether there are any differences in global narrative structure – as measured by the index score – attributable to grade, sex or experience.

5.3.2.1 Overall results

Fig. 5.22 presents the mean index score in first and fourth grade as well as its increase from first to fourth grade; Tab. 5.16 gives the descriptive statistics for the narrative index results in both grades.

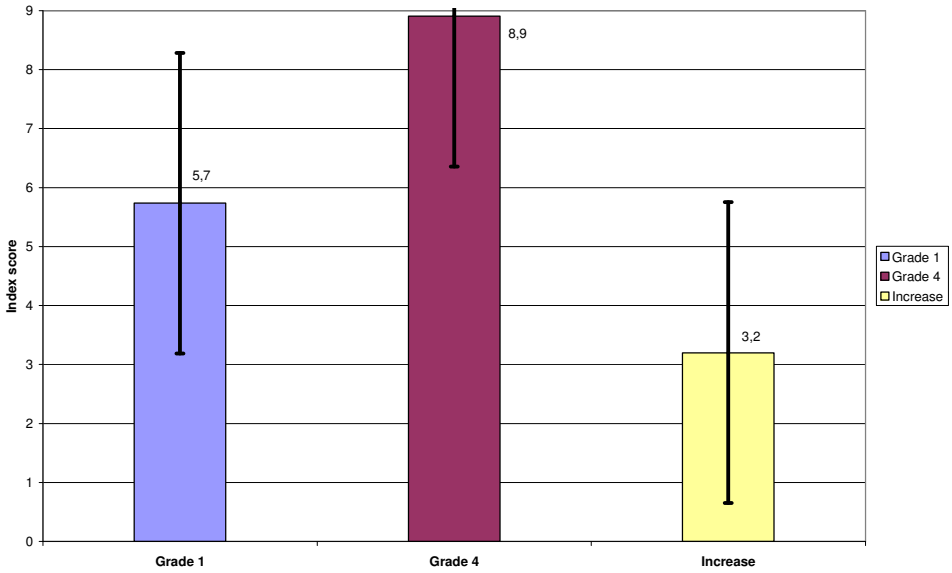

Fig. 5.22 Mean narrative index score by grade as well as increase from first to fourth grade

	Total N	Mean	Standard Deviation	Median	Mode	Minimum	Maximum
Grade 1	34	5,7	2,6	6	7	2	9
Grade 4	32	8,9	,4	9	9	7	9
Total	66	7,3	2,4	9	9	2	9

Tab. 5.16 Descriptive statistics narrative index by grade

First graders scored an average of 5,7 points on the index. By the end of fourth grade the mean index score has risen to 8,9, which basically corresponds to a full score. Statistically, this increase of the mean score from first to fourth grade is very highly significant and the effect size corroborates the strong influence of grade (3,2 points; $U=140$, $p<0{,}001$, $\eta^2=0{,}51$).[22] If one looks at the mean index score as a function of the maximal score of 9 – which would indicate a fully developed global narrative structure – first graders thus realize on average 64% of a global narrative structure, while for fourth graders the development of a global narrative structure is completed (99%).

Another difference between first and fourth grade concerns the interindividual variation, which is quite high in first grade as measured by a standard deviation of 2,6 points or 46% of the mean and a range of 7 points from lowest (2 points) to highest score (9). While the mean index score increases, however, the interindividual differences decrease to a standard deviation of merely 0,4 or 4% of the mean and a

[22] Grade as a single factor explains 51% of the variation in results.

range of 2 in fourth grade (cf. Tab. 5.16). This larger homogeneity in fourth grade is also reflected in Fig. 5.23, which presents the distribution of scores in first and fourth grade; Tab. 5.17 gives the corresponding values.

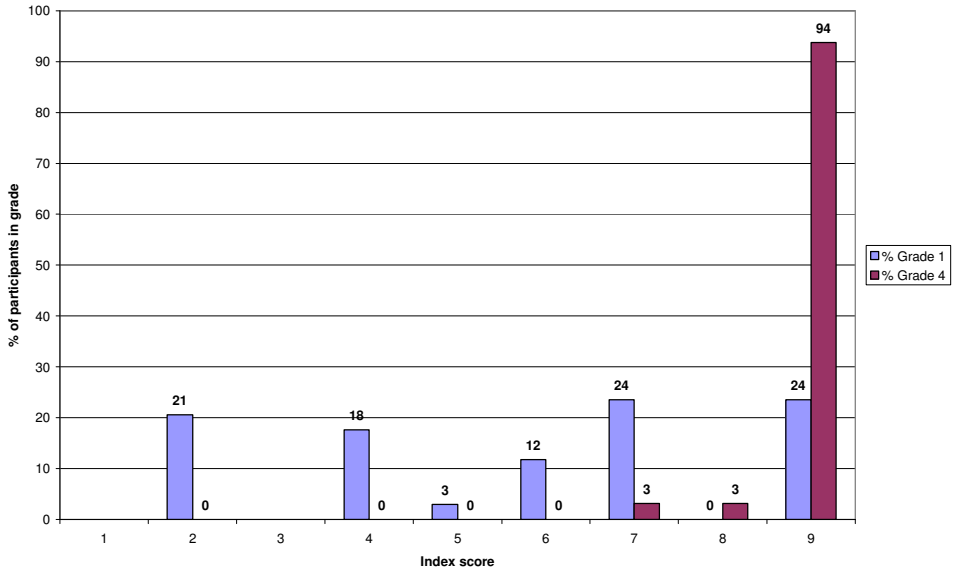

Fig. 5.23 Distribution of narrative index scores by grade

		Grade 1		Grade 4	
		N	% within grade	N	% within grade
Narrative index score	2	7	21%	0	0%
	4	6	18%	0	0%
	5	1	3%	0	0%
	6	4	12%	0	0%
	7	8	24%	1	3%
	8	0	0%	1	3%
	9	8	24%	30	94%
Total		34	100%	32	100%

Tab. 5.17 Distribution of narrative index scores by grade

The distribution shows that first graders' scores are distributed relatively evenly on a range from 2 to 9 with a little over half of them (53%, 18 of 34) scoring 2, 4, 5 or 6 points on the index and a little less than half of them (47%, 16 of 34) scoring 7 or 9 points.[23] Thus, 24% of first graders (8 participants) are already able to produce a fully globally organized narrative – as indicated by the maximum score of 9 points – and another 24% come close to it. In fourth grade 94% of participants reach the maximum index score and only two children produce lower scores, namely 7 (3%)

[23] Cf. also the median in first grade.

and 8 points (3%). That is, frequency distribution and mean score show that some first graders already perform like fourth graders with respect to a global structure but that a global organization of narratives has become the rule only in fourth grade. Since the present study only compares first and fourth grade data, it is, of course, not possible to draw any conclusion as to when exactly a global structure becomes predominant.

To sum up, the overall mean index score was found to increase strongly from first to fourth grade; the significance of this increase was confirmed by the statistical analysis. Thus, mean score and distribution showed that first graders' global organization is mostly still developing, even if there are large differences between them and roughly a quarter of first graders produced a fully developed narrative structure. It was also found that participants' interindividual differences decreased very strongly from first to fourth grade and that almost all fourth graders produced the highest index score, which means that the development of a global narrative structure can be seen as concluded by the end of fourth grade.

5.3.2.2 Index by sex

The mean index score by sex and grade as well as its increase are given in Fig. 5.24; the corresponding descriptive statistics are given in Tab. 5.18.

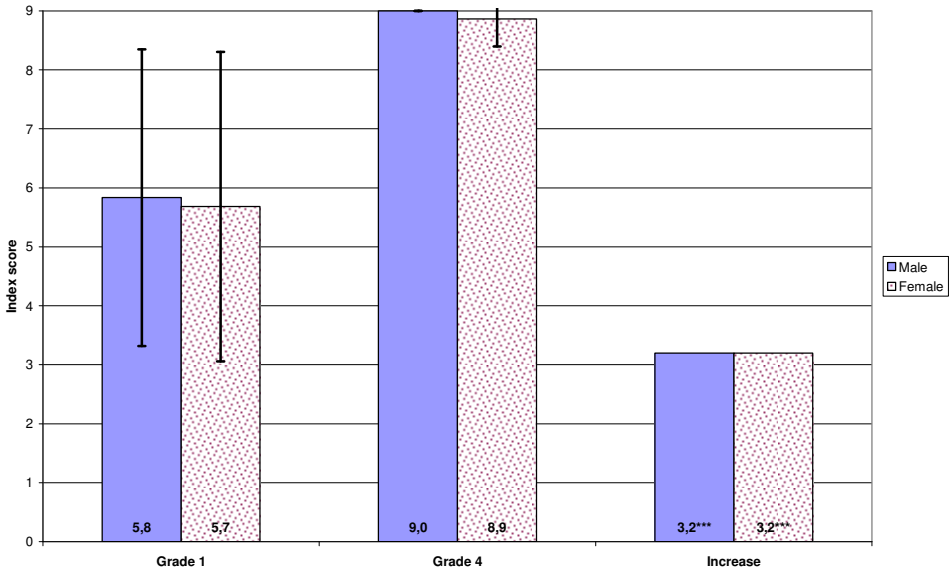

Fig. 5.24 Mean narrative index score by grade and sex as well as increase from first to fourth grade

		Narrative index						
		Total N	Mean	Standard Deviation	Median	Mode	Minimum	Maximum
Grade 1	Male	12	5,8	2,5	6	9	2	9
	Female	22	5,7	2,6	6,5	7	2	9
	Total	34	5,7	2,6	6	7	2	9
Grade 4	Male	10	9,0	0	9	9	9	9
	Female	22	8,9	,5	9	9	7	9
	Total	32	8,9	,4	9	9	7	9

Tab. 5.18 Descriptive statistics narrative index by sex and grade

Male and female first graders perform very similarly. Thus, there is neither a significant observable nor a statistically significant difference between the two groups' mean score in first grade (5,8 vs. 5,7; U=129,5, ns). Great similarity also exists in the interindividual variation as measured by the standard deviation, namely 2,5 versus 2,6 (43% and 46% in relation to the mean). Lowest (2 points) and highest score (9 points) as well as scoring range (7 points) are even identical.

The frequency distribution, which is shown in Fig. 5.25 (the corresponding statistics are given in Tab. 5.19), underlines this similarity between male and female first graders: Both groups' scores are very similarly distributed and differences can be considered negligible.

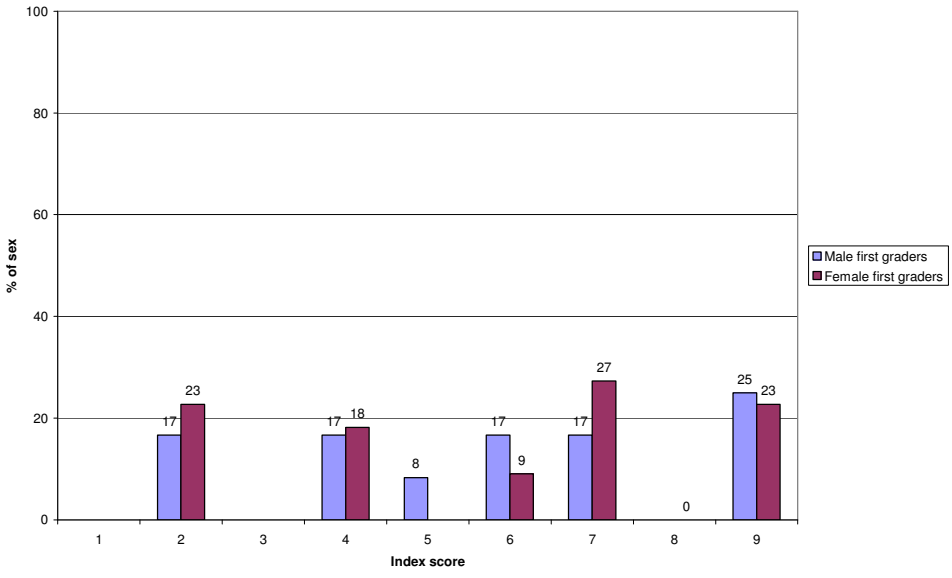

Fig. 5.25 Distribution of narrative index scores by sex in first grade

		Grade 1				Grade 4			
		Male		Female		Male		Female	
		N	% within grade	N	% within grade	N	% within grade	N	% within grade
Narrative index score	2	2	17	5	23	0	0	0	0
	4	2	17	4	18	0	0	0	0
	5	1	8	0	0	0	0	0	0
	6	2	17	2	9	0	0	0	0
	7	2	17	6	27	0	0	1	5
	8	0	0	0	0	0	0	1	5
	9	3	25	5	23	10	100	20	91
Total		12	100	22	100	10	100	22	100

Tab. 5.19 Distribution of narrative index scores by sex and grade

Both sexes' performance follows the general trend of a significant increase of the mean index score as well as a decrease of the interindividual variation (as measured by the standard deviation) from first to fourth grade (

Fig. 5.24 and Tab. 5.18). Both groups' mean score rises by approximately 50% and this increase is statistically (very highly) significant for both male ($U=15$, $p<0,05$, $\eta^2=0,53$) and female ($U=63$, $p<0,001$, $\eta^2=0,51$) participants. At the same time the standard deviation of male participants decreases from 2,5 to 0, i.e. from 43% of the mean to 0%, while the one of the female participants decreases from 2,6 to 0,5, i.e. from 46% to 6% of the mean (Tab. 5.18).

In fourth grade the performance of male and female participants is even more similar. Thus, there is again neither a significant observable nor a statistically significant difference between the two groups' mean score (9,0 vs. 8,9; $U=100$, ns) and great similarity exists in the interindividual variation as measured by the standard deviation (0 vs. 0,5). Even median and mode are identical (both 9 points) for both groups. The only observable difference between female and male fourth graders is their minimum and maximum score, i.e. their range of scores (2 and 0, respectively). This is reflected in Fig. 5.26, which shows the frequency distribution for male and female fourth graders' index score.

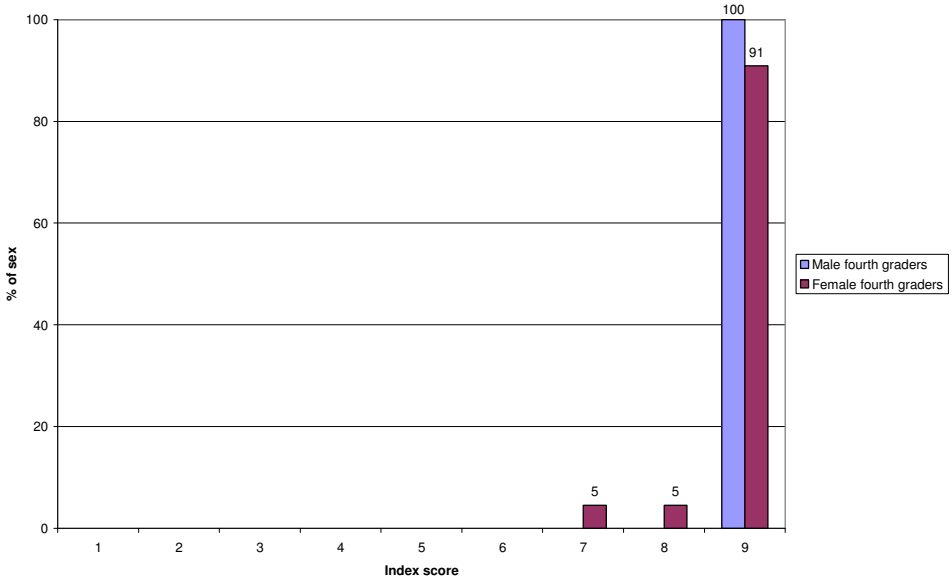

Fig. 5.26 Distribution of narrative index scores by sex in fourth grade

The distribution shows that the only two fourth graders scoring fewer than 9 points are both females. Male fourth graders thus seem to perform slightly better, since they all obtain 9 points on the index. In absolute numbers, however, this means that two of the females did not score at ceiling. Again, keeping in mind the large difference in group size, this “male advantage” seems to be attributable rather to a group size effect than a true difference between male and female participants.

To sum up, both sexes followed the general trend of an increase in the mean index score as well as a decrease in interindividual variation. At the same time neither significant observable nor statistical differences between male and female participants were found for either grade.

5.3.2.3 Index by experience group

The mean index scores obtained in first and fourth grade by the three different experience groups as well as the increase from first to fourth grade are shown in Fig. 5.27; the corresponding descriptive statistics are given in Tab. 5.20. As in previous sections a main comparison will be made between participants with bilingual preschool experience (bilis) and those without (monos); the small MB group will be considered separately.[24]

[24] As explained earlier the MB group is very small in size and at the same time very heterogeneous in all of its results so that it cannot be considered representative.

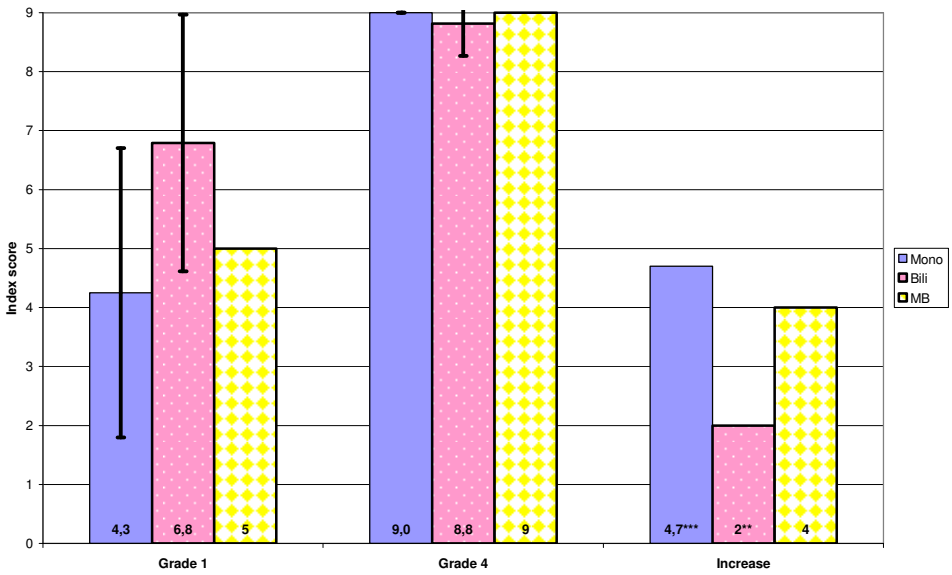

Fig. 5.27 Mean narrtive index score by experience group and grade as well as increase from grade 1 to 4. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

		Narrative index						
		Total N	Mean	Standard Deviation	Median	Mode	Minimum	Maximum
Grade 1	Mono	12	4,3	2,5	4	2	2	9
	Bili	19	6,8	2,2	7	9	2	9
	MB	3	5,0	-	-	2	2	7
	Total	34	5,7	2,6	6	7	2	9
Grade 4	Mono	12	9,0	0	9	9	9	9
	Bili	16	8,8	0,5	9	9	7	9
	MB	4	9,0	-	-	9	9	9
	Total	32	8,9	0,4	9	9	7	9

Tab. 5.20 Descriptive statistics narrative index by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

First graders with bilingual preschool experience clearly outperform those without such prior experience: The bilingual's mean index score is 6,8 as opposed to 4,3 for the monos. Statistically, this difference is also very highly significant with a solid effect size ($U=50,5$, $p<0,01$, $\eta^2=0,23$). In terms of a fully global narrative structure (indicated by nine points on the index) this means that bili first graders realize on average 76% of a global narrative structure, while the mono group's result corresponds to merely 48% of a global narrative structure. At the same time the mono group is much more heterogeneous in first grade with a standard deviation of

2,5 or 60% in relation to the mean as opposed to 2,2 but only 32% of the mean in the bili group.

Thus, bili first graders already realize large parts of a global narrative structure. Mono first graders, on the other hand, realize some of the important components but do not yet seem to use them in their function for a global narrative organization. This finding confirms the results obtained for the total number of components (cf. ch. 5.1.1.3). At the same time the interindividual differences in first grade are very large for both experience groups just as they were very large for the total number of components; the same is true for the MB group as will be described further below.

Bili first graders' advantage is corroborated by median and mode (cf. Tab. 5.20) as well as the frequency distribution in first grade; all of these are illustrated in Fig. 5.28, which shows the frequency distribution of the three experience groups in first grade. The corresponding values are given in Tab. 5.21.

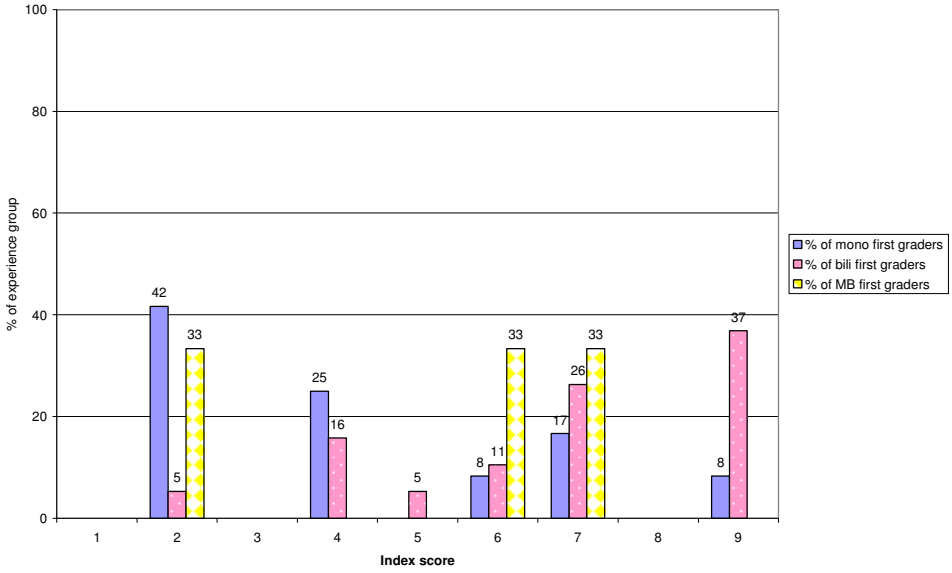

Fig. 5.28 Distribution of narrative index scores by experience group in first grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

Narrative index score	Grade 1 (N=34)						Grade 4 (N=32)					
	Mono		Bili		MB		Mono		Bili		MB	
	N	% within grade	N	% within grade	N	% within grade	N	% within grade	N	% within grade	N	% within grade
2	5	42	1	5	1	33	0	0	0	0	0	0
4	3	25	3	16	0	0	0	0	0	0	0	0
5	0	0	1	5	0	0	0	0	0	0	0	0
6	1	8	2	11	1	33	0	0	0	0	0	0
7	2	17	5	26	1	33	0	0	1	6	0	0
8	0	0	0	0	0	0	0	0	1	6	0	0
9	1	8	7	37	0	0	12	100	14	88	4	100
Total	12	100	19	100	3	100	12	100	16	100	4	100

Tab. 5.21 Distribution of narrative index scores by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

The distribution demonstrates that over one third of bili first graders (37%, i.e. 7 of 19) scored the highest possible number of points on the index, namely nine, as opposed to only one child of the mono group (8%, 1 of 12). The two highest scores taken together, i.e. seven and nine points, are obtained by the majority of bili first graders, namely 63% (12 of 19) as compared to only 25% of mono first graders (3 of 12). Almost half of the mono group (42%, 5 of 12), on the other hand, scored the lowest number of points observed, namely two, as opposed to only one bili first grader (5%, 1 of 19). At the same time the two lowest scores together (two and four points) account for the majority of mono first graders' results, namely 67% (8 of 12), but for only 21% of the bilis (4 of 19). Consequently, an index score between 2 and 4, i.e. a largely undeveloped global structure, seems to be more typical for mono first graders, while a score between 7 and 9, i.e. an (almost) fully developed global organization, seems to be more typical for bili first graders. Thus, the distribution of index scores confirms the conclusions drawn from the mean index scores.

MB first graders form an intermediate group: Their mean index score (5 points) is a little higher than that of the mono group (4,3) but much lower than the bili group's mean score (6,8). At the same time the MB group's results are very heterogeneous, even though the interindividual variation as measured by the scoring range is the lowest of all three experience groups. More specifically, the three MB first graders scored 2 (G1-C1-11), 6 (G1-C1-18) and 7 points on the index (G1-C1-5). That is, no overall pattern emerges.

All three experience groups follow the general trend of an increase in mean score and a decrease in interindividual variation from first to fourth grade (Fig. 5.27 and Tab. 5.20). The mono group's mean index score increased most strongly (+4,7); statistically this increase was also very highly significant (U=6, p<0,001, η^2=0,76). The second highest increase was obtained by the MB group (+4).[25] The bili group had the comparatively lowest increase (+2), which was, nevertheless, statistically highly significant (U=65,5, p<0,01, η^2=0,31). These differences in increase can be explained by the first grade scores: Due to their much lower score in first grade, mono and MB group have a larger necessity and additionally more room for increase as opposed to the bili group, which already scored high in first grade. This is also reflected in the effect sizes of bili and mono group, since grade explains 31% of the variation in bili but 76% in mono index scores.

As a result of the different increase rates all three experience groups perform very similarly in fourth grade (Fig. 5.27 and Tab. 5.20). The mean score of mono and MB group is at ceiling (9 points on the index), which means that all of them realize a fully global narrative structure. The bili group, however, has a slightly lower mean score of 8,8 points, which corresponds to 98% of a global narrative structure. This lower mean score is explained by the bili fourth graders' interindividual difference. That is, bili participants are the only experience group with any variation in fourth grade (SD=0,5 or 6% of the mean); their variation is reflected in the frequency distribution in fourth grade, which is given in Fig. 5.29.

[25] The MB group was not submitted to a statistical analysis due to its small number of members (cf. ch. 4.1.2).

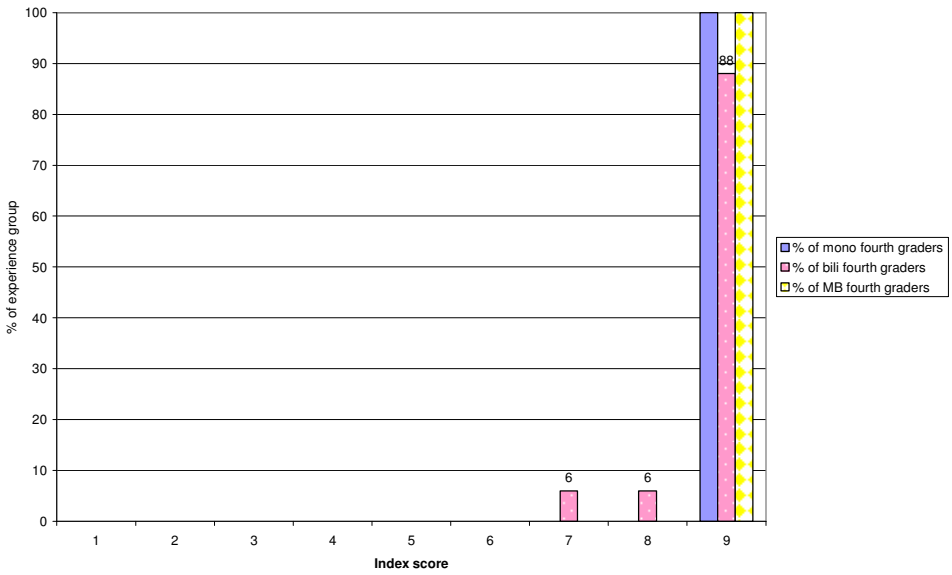

Fig. 5.29 Distribution of narrative index scores by experience in fourth grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

As the distribution shows, 100% of monos and MBs score 9 points on the index as opposed to only 88% of the bilis. Thus, 6% of bilis score seven, and 6% eight components. In absolute numbers, however, these 12% correspond to only two of the 16 bili fourth graders (G4-C1-15 and G4-C5-20)[26]. In combination with the fact that the difference between mono and bili fourth graders is not statistically significant (U=84, ns)[27] this leads to the conclusion that there is no significant difference in fourth grade results attributable to L2 preschool experience.

To sum up, L2 preschool experience was again found to have a significant impact on first but not fourth grade results: Mean score and distribution showed that the majority of bili first graders was able to realize either a fully global narrative structure or large parts of it, while this was not the case for either mono or MB group. At the same time the bili results were also most homogeneous. The index score of all three experience groups followed the overall trend of an increase in mean score and homogeneity of results. Due to different increase rates, however, there was no significant influence of L2 preschool experience in fourth grade. That is, all three experience groups told stories with a fully global narrative structure in fourth grade. The statistical analysis confirmed the significant bili advantage in first grade, the

[26] G4-C1-15, who scored 7 points, was not given credit for the SETTING since the dog was not mentioned. G4-C5-20 did not produce a second attempt and thus scored 8 points.
[27] The MB group was excluded from the statistical analysis due to its small size.

significant increase in mean score from first to fourth grade as well as the lack of significant differences in fourth grade.

5.3.3 Summary: Narrative index

This section presented the results for a narrative index indicating the degree of coherence of participants' stories, i.e. the degree of a global narrative structure, on a range from 1 to 9 points. The purpose of this index was to complement the results obtained for both total number of components and individual narrative components by focusing on a smaller selection of components crucial for producing a coherent story. Thus, it was investigated whether there were any differences in global narrative structure – as measured by the index score – attributable to grade, sex or experience.

Grade was found to have a significant influence with respect to mean index score and interindividual differences: Observed and statistical results showed that fourth graders obtained a significantly higher mean score than first graders (5,7 vs. 8,9). Participants' interindividual variation, on the other hand, was significantly lower in fourth grade. At the same time a remarkable overlap between first and fourth grade mean scores was observed, which showed that quite a large number of first graders was able to produce an (almost) fully global narrative structure. That is, 48% of first graders (N=16) achieved a mean index score obtained also by at least one of the fourth graders and half of these first graders (24%, N=8) even realized the maximum score of 9 points on the index.

Sex did not significantly influence participants' index results. Thus, both sexes followed the general trend of an increase in the mean score and a decrease in interindividual variation from first to fourth grade. At the same time neither significant observable nor statistical differences between male and female participants were found in either grade. L2 preschool experience, on the other hand, had a strong influence on participants' index score. However, this influence was limited to first grade. That is, the results by experience group qualified the overall results in that an (almost) fully global organization was produced mainly by first graders with L2 preschool experience. That is, seven of the eight first graders (88%) achieving 9 points and five of the eight first graders (63%) achieving 7 points belonged to the bili group. Merely one mono first grader, on the other hand, obtained 9 points and only two of them 7 points. One MB first grader also achieved 7 points.

To sum up, the narrative index results showed that first as well as fourth graders were able to produce fully coherent narratives. However, almost exclusively first graders with bilingual preschool experience were able to produce fully globally organized stories, i.e. an index score of 9 points. It is only by the end of fourth grade that a global organization has become the rule for all participants.

5.4 Narrative coherence: Summary

In the previous sections the results of three different measures were described, which were used to investigate the global narrative structure of participants' stories, i.e. their coherence: The total number of components in participants' stories, the frequency of each of the 14 narrative components under investigation, and an index of narrative structure consisting of a reduced number of components indicating a globally organized story. All results were examined as to a possible influence of grade, sex, and L2 preschool experience.

Based on the findings of previous studies (cf. ch. 3) the following hypotheses had been put forward with respect to the research questions addressed in the analysis of narrative coherence (cf. ch. 2.3):

(1) Grade/age has a significant influence on the narrative coherence of participants' stories
(2) Participants' stories become more coherent from first to fourth grade as measured by the number of narrative components
(3) There are qualitative differences in narrative coherence between grades as measured by differences in frequency among the individual narrative components.
(4) Sex does not have a significant influence on the narrative coherence of participants' stories.

First and foremost it was found that all first graders were able to produce at least one narrative component (minimum score) and that none of the first or fourth graders realized all 14 components; the highest score was 12. At the same time each of the 14 narrative components was produced by at least one participant in first as well as in fourth grade.

All three measures were strongly influenced by grade. That is, participants produced significantly more and a greater variety of components in fourth than in first grade, every narrative component was produced more often by fourth graders, and

participants' index scores were significantly higher in fourth grade. Hypotheses (1), (2) and (3) were thus confirmed. At the same time the interindividual variation – measureable by total number of components and index score – was significantly lower in fourth than in first grade. More qualitatively, narrative index and an order of difficulty based on the frequency of the individual components showed that the coherence of participants' stories is only beginning to emerge in first grade. By the end of fourth grade, however, it is fully developed for almost all participants. Consequently, a developmental pattern emerges: The coherence of participants' stories strongly increases from first to fourth grade, and by the end of fourth grade the development of coherence can be seen as concluded. However, a closer look at the distributions of total number of components and index score qualified this overall pattern, since there was a large overlap between first and fourth grade scores. That is, almost half of all first graders (48%) produced stories which were as coherent as those of one or more fourth graders and 24% of first graders told fully coherent stories (as measured by an index score of 9 points).

Participants' sex did not have any significant influence on coherence results. That is, male and female participants performed very similarly in first and fourth grade regarding all three measures of coherence. Additionally, they followed the overall developmental pattern described above. Hypothesis (4) was thus also confirmed.

L2 preschool experience, on the other hand, was found to have a significant influence on all three coherence measures in first grade, which qualifies the overall results. That is, first graders with L2 preschool experience (the 'bili' group) produced a significantly higher number of components, they realized almost every component more often, and they obtained a significantly higher index score than first graders without prior L2 experience. At the same time bili first graders' results showed less interindividual variation. This means that the overlap of results between first and fourth grade described above is largely attributable to the first grade results of children with L2 preschool experience. Thus, first graders with bilingual preschool experience produced significantly more often coherent stories and a very substantial number of them[28] produced as (fully) coherent stories as fourth graders. Due to distinct increase rates, however, these differences even out until the end of fourth

[28] 63% as measured by the index and 69% as measured by the total number of components and thus the majority of bili first graders. If "untypical" fourth grade scores are disregarded, however, the percentages decrease to 37% and 32%, respectively.

grade so that by then coherent stories have become the rule for all participants regardless of experience group.

6 The development of cohesion: Results

6.1 Overall cohesive density

This section aims at answering the question how cohesive participants' stories are as measured by their overall degree of cohesive density and whether there are any differences in cohesion attributable to grade, sex or L2 preschool experience. The observed results will be described first, followed by the statistical results. Finally, the results obtained will be summarized.

As stated in ch. 4 and restated at the beginning of ch. 5, possible differences between the longitudinal and cross-sectional data sets were explored for all overall measures of cohesion (and coherence) as well as for randomly selected individual variables. No significant differences were found. Thus, the respective first and fourth grade cohorts were treated as one first/fourth grade data set and cohorts are not differentiated in the discussion of any of the cohesion results in this section or the following sections.

6.1.1 Observed results overall cohesive density

Participants' mean cohesive density in first and fourth grade as well as the corresponding increase is depicted in Fig. 6.1; the descriptive statistics are given in Tab. 6.1.

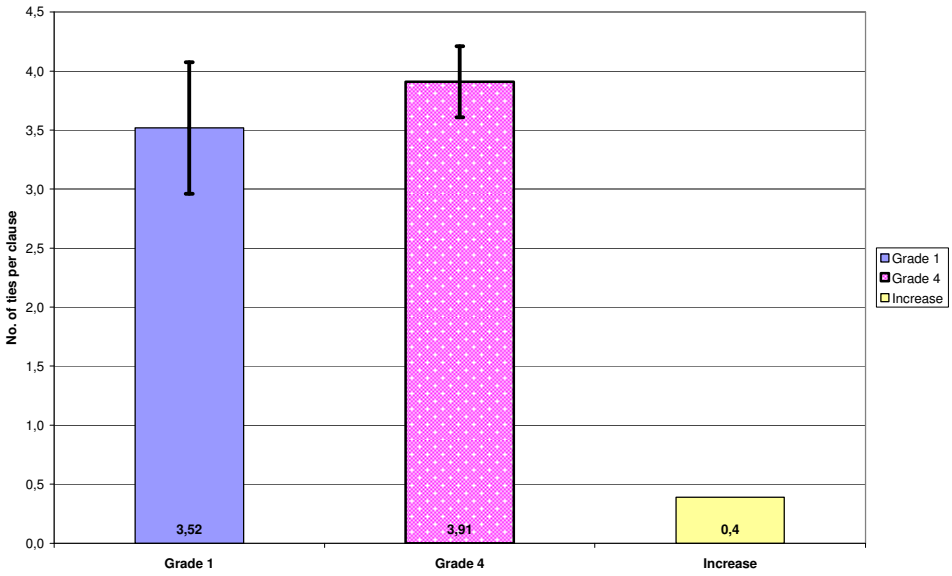

Fig. 6.1 Mean cohesive density by grade as well as increase from first to fourth grade

		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Cohesive density	Grade 1	34	3,52	,56	3,47	2,17	4,44
	Grade 4	32	3,91	,30	3,89	3,43	4,65
	Total	66	3,71	,49	3,75	2,17	4,65

Tab. 6.1 Descriptive statistics cohesive density by grade

All first graders were able to produce at least some cohesive ties; this is indicated by the first grade minimum score of 2,17 ties per clause. On average first graders realized 3,52 ties per clause. That is, first graders produced an average of 3,52 cohesive devices per clause – or approximately seven in two clauses – which refer to a preceding textual element.[1] At the same time the interindividual variation as measured by the standard deviation is low compared to the ones found for coherence, namely 0,56 or only 16% in relation to the mean (variability ratio). The range from lowest (2,17) to highest score (4,44) is quite high, on the other hand, i.e. the first grader with the highest cohesive density produces almost twice as many cohesive devices per clause than the first grader with the lowest score.

From first to fourth grade participants' overall cohesive density increases by 0,39, i.e. fourth graders produce roughly 4 cohesive devices more in 10 clauses. As a result, fourth graders averaged a cohesive density of 3,91. That is, fourth graders produced roughly four cohesive ties per clause. At the same time the interindividual

[1] In the following, the terms 'cohesive device' and '(cohesive) tie' will be used interchangeably since only cohesive devices forming cohesive ties were considered for the analysis.

variation decreases from 0,56 or 16% of the mean in first grade to an even lower 0,3 or 8% of the mean in fourth grade. This further reduction of the interindividual variation is also evident from the range of results, which decreases from 2,27 to 1,22.

Fig. 6.2 presents the distribution of participants' individual cohesive density scores in first and fourth grade; the corresponding values are given in Tab. 10.1 and Tab. 10.2 in the Appendix (ch. 10.1) .

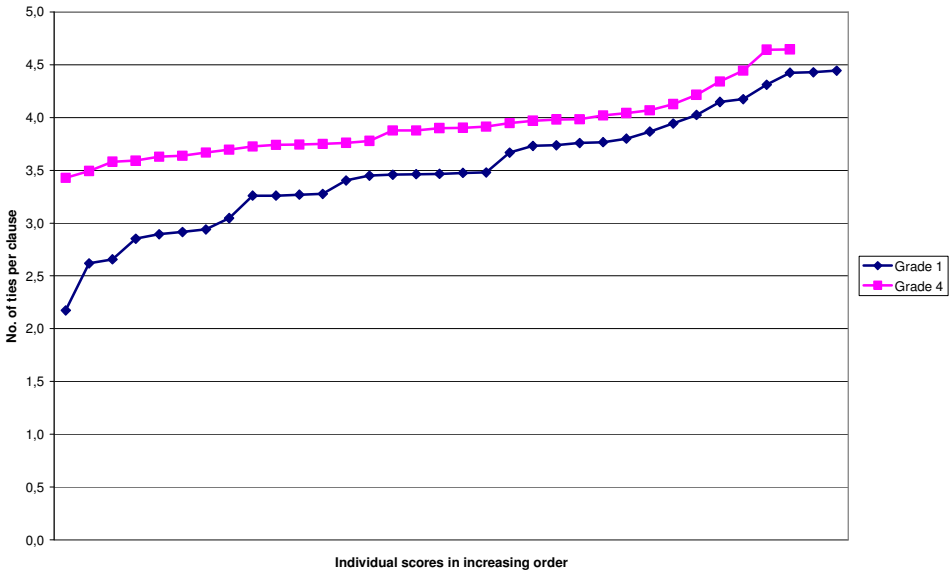

Fig. 6.2 Distribution of participants' individual cohesive density scores by grade

As Fig. 6.2 shows, the first grade curve is much steeper than the one in fourth grade, which reflects the higher degree of variation. At the same time the distribution qualifies the observed difference in mean cohesive density between first and fourth grade. That is, there is an overlap of scores between 3,43 (fourth grade minimum) and 4,44 ties per clause (first grade maximum). 94% of fourth graders (30 of 32) as compared to 62% of first graders (21 of 34) obtained a cohesive density within this range. At the same time scores below 3,43 are realized exclusively by first graders (38%, 13 of 34) and the only two participants (6%, G4-C5-12 and G4-C5-18) with a higher cohesive density than the overlap are fourth graders. Thus, the distribution indicates that the main difference between first and fourth grade cohesive density lies in the 38% of first graders producing cohesive density scores below the overlap. An additional comparison of minimum and maximum scores in first and fourth grade shows that the minimum score increases much more strongly (by 1,26) than the maximum score (by 0,21). This leads to the conclusion that changes in cohesive density from first to fourth grade mainly concern an increase in the lower scoring ranges, while there is a comparative stability in the higher ranges.

To sum up, fourth graders' stories were found to have a higher mean cohesive density and a smaller amount of interindividual variation, even if the results were comparatively homogeneous in both grades. However, the distribution of results showed that more than half of all first graders produced a cohesive density that could also have been produced by a fourth grader. At the same time the distribution indicated that the major change in cohesive density from first to fourth grade concerns the lower density scores, which increase, while the higher scoring ranges largely remain the same.

6.1.2 Overall cohesive density by sex

Male and female participants' mean cohesive density in first and fourth grade as well as the corresponding increase are depicted in Fig. 6.3; the descriptive statistics are given in Tab. 6.2.

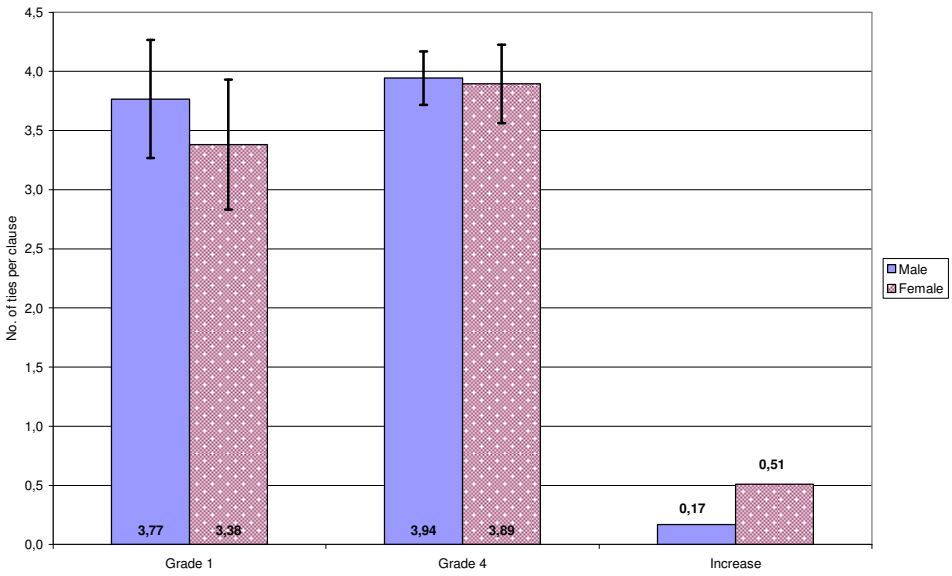

Fig. 6.3 Mean cohesive density by grade as well as increase from first to fourth grade

		Cohesive density (No. of ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Male	12	3,77	,50	3,70	2,85	4,44
	Female	22	3,38	,55	3,43	2,17	4,42
	Total	34	3,52	,56	3,47	2,17	4,44
Grade 4	Male	10	3,94	,23	3,94	3,64	4,34
	Female	22	3,89	,33	3,88	3,43	4,65
	Total	32	3,91	,30	3,89	3,43	4,65

Tab. 6.2 Descriptive statistics cohesive density by grade and sex

In first grade male participants perform (M=3,77) better than female participants (M=3,38), since they produce an average of 0,39 ties more per clause or 4 in 10 clauses. However, the median indicates that 50% of male participants actually score 3,7 or higher, while 50% of females score 3,43 or higher, i.e. the median points to the difference between sexes being smaller than what is indicated by the mean. At the same time male and female first graders' interindividual variation is very similar and – just as for the overall results – comparatively low. Thus, male results have a standard deviation of 0,5 or 13% of the mean, while females show a marginally higher variation of 0,55 or 16% of the mean. The same is evident in the range of scores which is 1,59 for male first graders and 2,25 for female first graders; the higher scoring range for female first graders is most likely attributable to a lower minimum score and their larger group size.

The distribution of participants' individual cohesive density scores in first grade as a function of sex is depicted in Fig. 6.4; the corresponding values are given in Tab. 10.3 and Tab. 10.4 in the Appendix (ch. 10.2).

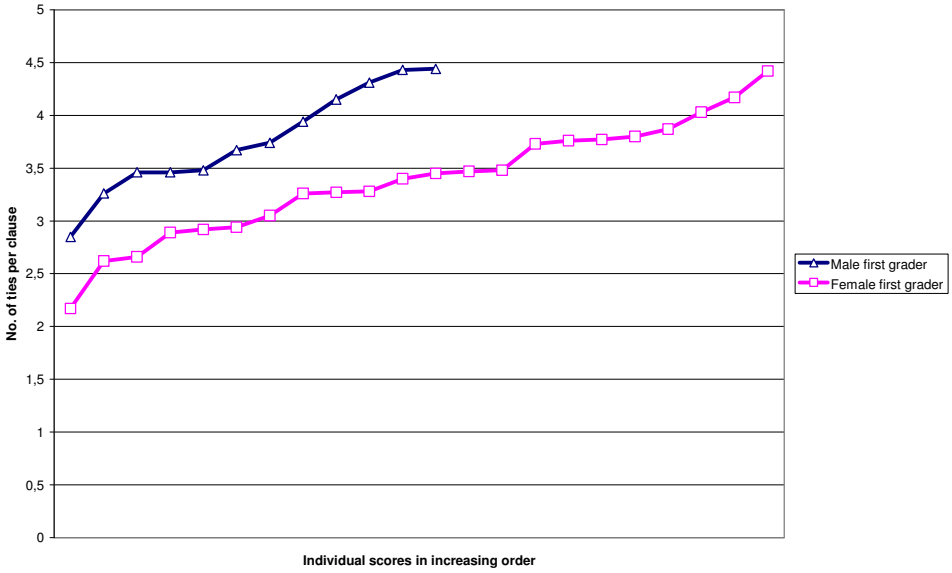

Fig. 6.4 Distribution of participants' individual cohesive density scores by sex in first grade

The curves confirm the slightly larger interindividual variation in female first graders' results described above. However, it shows again a great similarity in results in terms of an overlap between male and female first graders' cohesive density scores. This overlap ranges from the male minimum of 2,85 to the female maximum of 4,42 and it includes 83% of male first graders (10 of 12) as well as 86% of female first graders (19 of 22). Three female first graders (14%) produced a cohesive density below the overlap,[2] while two male first graders (17%, C1-G1-11 and C1-G1-2), scored marginally above (4,43 and 4,44, respectively). However, these differences seem negligible since the number of participants in each group is so different (male N=12, female N=22). From the large overlap and the relatively few participants scoring below or above the overlap it would rather seem that the slight male advantage identified above is attributable to a group size effect.

Both male and female mean density increase from first to fourth grade. However, female participants' cohesive density increases by 0,51, i.e. by around 5 cohesive devices per 10 clauses, while males' cohesive density increases only one third of that, namely by 0,17, i.e. close to two cohesive devices per 10 clauses. As a result, male and female fourth graders' results are even more similar (Fig. 6.3, Tab. 6.2); the difference between male (M=3,94) and female (M=3,89) mean cohesive density is now only 0,05, i.e. 5 cohesive devices per 100 clauses.

[2] C1-G1-4 achieved a cohesive density of 2,17, C8-G1-18 scored 2,62, and C8-G1-21 scored 2,66.

As the mean score increases, the interindividual variation becomes even smaller. Thus, the standard deviation for male participants decreases from 0,5 or 13% of the mean in first to 0,23 or merely 6% in fourth grade, while for females it decreases from 0,55 or 16% of the mean to 0,33 or 8%. The same is indicated by the range of scores, which drops from 1,59 in first to 0,7 in fourth grade for male participants and from 2,25 to 1,22 for female participants. Standard deviation and range indicate that female participants continue to show a slightly larger variation in results in fourth grade. Altogether, however, these differences are as much negligible as in first grade – especially when keeping in mind the large difference in group size.

Fig. 6.5 presents the distribution of participants' individual cohesive density scores in fourth grade as a function of sex; the corresponding values are given in Tab. 10.5 and Tab. 10.6 in the Appendix (ch. 10.2).

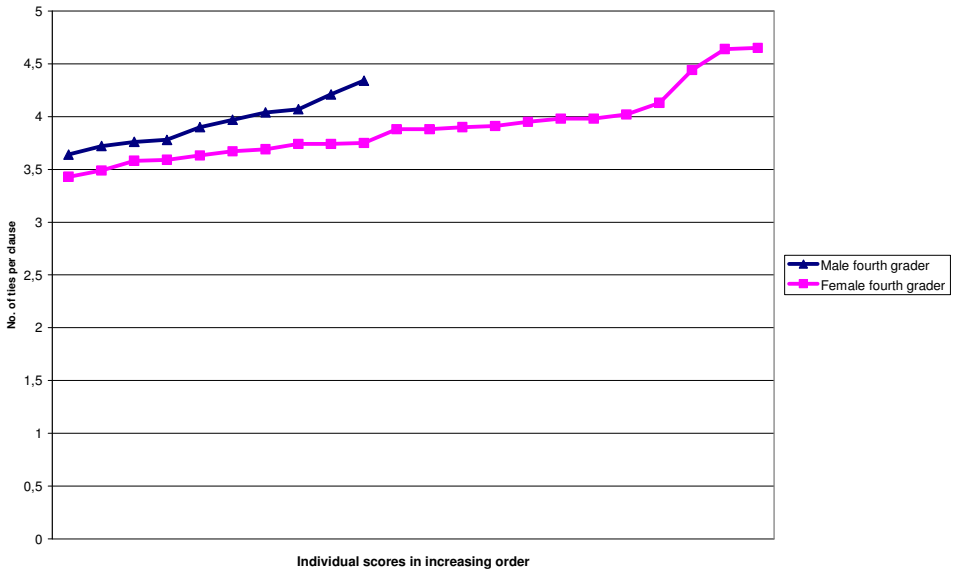

Fig. 6.5 Distribution of participants' individual cohesive density scores by sex in fourth grade

Again, there is a strong overlap between male and female cohesive density. Thus, all male (10, i.e. 100%) and 64% of female fourth graders (14 of 22) achieved a cohesive density between 3,64 (male fourth grade minimum) and 4,34 ties per clause (male fourth grade maximum). 23% of the remaining female fourth graders (5 of 22) scored below that range,[3] while three female first graders (14%) scored above the overlap.[4] However, any male advantage in fourth grade distribution is again more likely attributable to a group size effect than to participants' sex.

[3] Up to 0,21 ties per clause or roughly 2 in 10 clauses less.
[4] Up to 0,31 ties per clause or about 3 ties more per 10 clauses

Fig. 6.6 shows the distribution of male and female individual cohesive density scores in first and fourth grade.

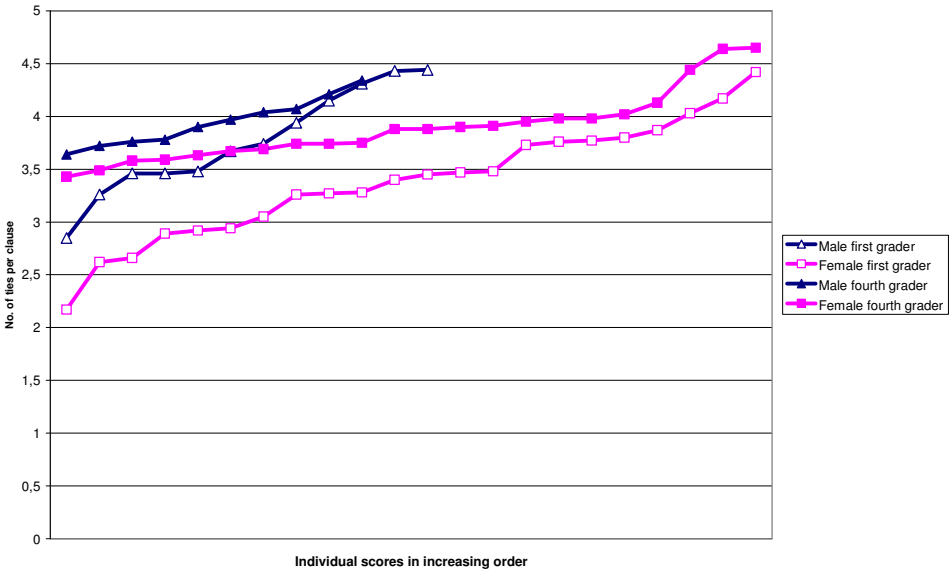

Fig. 6.6 Distribution of participants' individual cohesive density scores by sex and grade

A comparison of both sexes' first and fourth grade curves reflects first of all an increase in each sex's minimum cohesive density from first to fourth grade (cf. also Tab. 6.2). At the same time the curves show that there are large overlaps between first and fourth grade for both sexes. Male participants' cohesive density overlaps in the range of 3,64 (fourth grade minimum) to 4,34 ties per clause (fourth grade maximum). All of the 10 fourth graders (100%) score within this range as opposed to only 42% of first graders (5 of 12). 42% of the remaining male first graders (5 of 12) obtained a cohesive density below the overlap, while two of them (17%) scored above it. Female participants' cohesive density, on the other hand, overlaps between 3,43 (fourth grade minimum) and 4,42 ties per clause (first grade maximum). 86% of female fourth graders (19 of 22) but only 50% of first graders (11 of 22) scored in this range. The remaining 50% of female first graders (11) scored below the overlap, while the remaining three fourth graders (14%) scored above it. Thus, a comparison of the first and fourth grade curves for both sexes confirms the general trend outlined above, namely a reduction in lower scores with a substantial overlap between first and fourth graders in the upper scoring ranges.

To sum up, only negligible differences in cohesive density were observed between male and female participants in either grade and these differences may be attributable to a strong difference in group size. At the same time the analysis

confirmed the pattern identified for the overall results: Each sex's mean score increased from first to fourth grade and the interindividual variation, which was already low in first grade, decreased. The distribution for male and female participants' cohesive density also showed that a large percentage of first graders – males as well as females – produced stories which were as cohesive as those of their fourth grade peers and that for both sexes the main difference between grades involved a disappearance of the lower first grade density scores.

6.1.3 Overall cohesive density by experience group

Participants' mean cohesive density by grade and experience group is given in Fig. 6.7 together with its increase from first to fourth grade; the corresponding descriptive statistics are given in Tab. 6.3. As in earlier chapters, a first main comparison will be made between participants with and without L2 preschool experience (mono vs. bili group), while the small group of children from a monolingual group in a mixed preschool (MB group) will be considered separately.

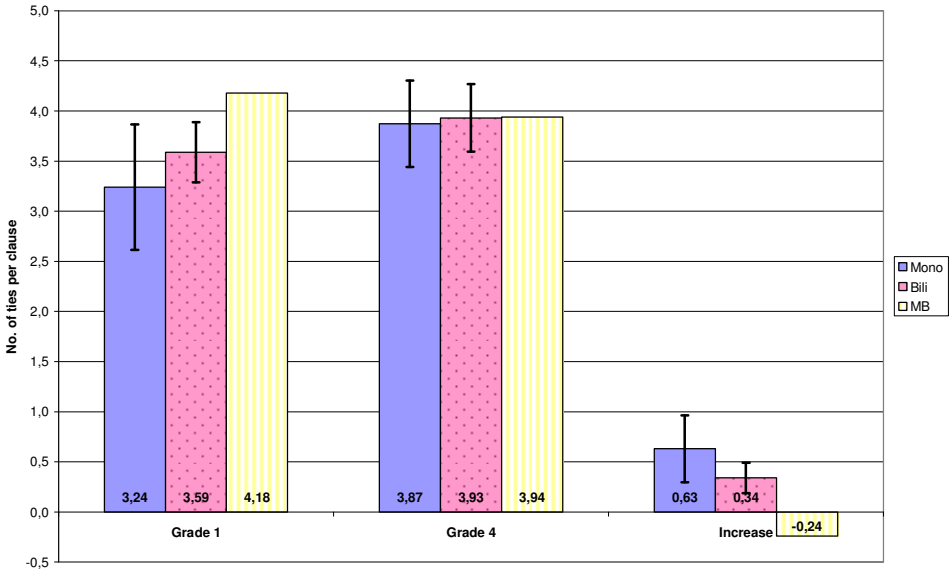

Fig. 6.7 Cohesive density by grade and experience group. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

		Cohesive density (No. of ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Mono	12	3,24	,63	3,15	2,17	4,42
	Bili	19	3,59	,43	3,67	2,66	4,44
	MB	3	4,18	-	-	3,80	4,43
	Total	34	3,52	,56	3,47	2,17	4,44
Grade 4	Mono	12	3,87	,30	3,83	3,58	4,65
	Bili	16	3,93	,34	3,89	3,43	4,64
	MB	4	3,94	-	-	3,74	4,07
	Total	32	3,91	,30	3,89	3,43	4,65

Tab. 6.3 Descriptive statistics cohesive density by grade and experience group. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

First graders with L2 preschool experience achieve a somewhat higher cohesive density than children without such prior experience. Thus, the bili group achieved an average of 3,59 ties per clause or roughly 36 per 10 clauses, while the mono group produced 3,24 ties per clause or roughly 32 per 10 clauses – a difference of 0,35 cohesive devices per clause or almost 4 in 10 clauses. As the median shows, mono and bili results are slightly skewed, however, and the bili group's advantage is a little more pronounced than the mean suggests. At the same time mono first graders' results are a little more heterogeneous than those of the bilis. This lower heterogeneity is evident from the standard deviation, which is 0,63 or 19% of the mean for the mono group as opposed to 0,43 or 12% for the bili group, and from the range of cohesive density scores, which is also higher for mono than bili first graders (2,25 vs. 1,78).

MB first graders (N=3) outperform both mono and bili first graders (Tab. 6.3) in that they obtain the highest mean score of all three experience groups, namely 4,18 ties per clause or approximately 42 ties in 10 clauses. At the same time they have the lowest variation in results as measured by their comparatively low scoring range (0,63) from a minimum of 3,8 to a maximum score of 4,42 ties per clause.

The distribution of mono and bili first graders' individual cohesive density scores is depicted in Fig. 6.8; the corresponding values are given in Tab. 10.7 and Tab. 10.8 in the Appendix (ch. 10.3).

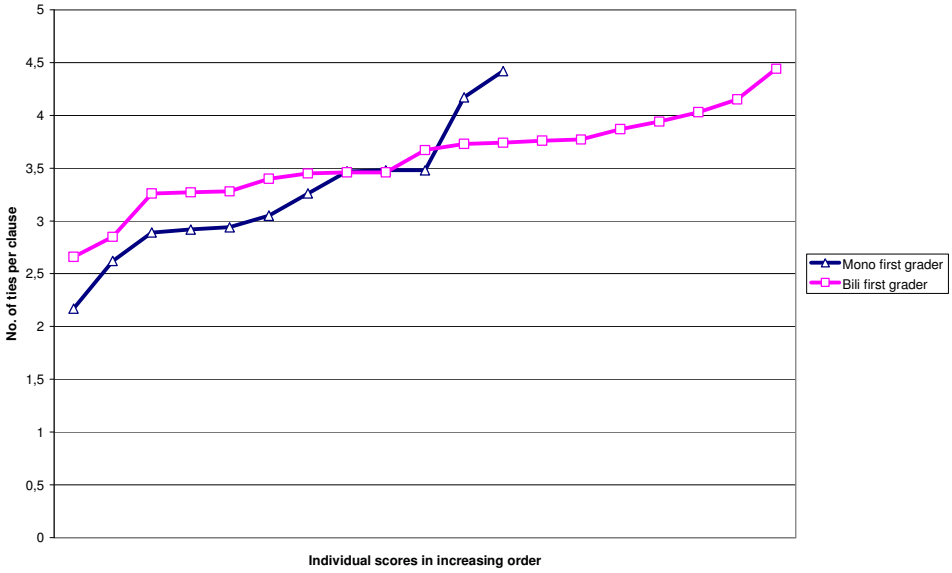

Fig. 6.8 Distribution of individual cohesive density scores by experience group in first grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

There is a strong overlap between mono and bili first graders, which ranges from 2,66 (bili minimum) to 4,42 ties per clause (mono maximum). 83% of mono (10 out of 12) and 95% of bili first graders (18 out of 19) score within this range. The two remaining mono first graders (17%, C1-G1-4, C8-G1-18) score (slightly) below this range (2,17 and 2,62, respectively). The remaining bili first grader (5%, C1-G1-2), on the other hand, scored slightly above the overlap and at the same time achieved the highest cohesive density of all participants, namely 4,44 ties per clause. Thus, the distribution of individual cohesive density results indicates again a marginal advantage of first graders with L2 preschool experience.

The distribution also points towards an influence of L2 preschool experience on the overall overlap between first and fourth graders described at the beginning of this chapter. Thus, the overlapping range of 3,43 (fourth grade minimum) to 4,44 (first grade maximum) was applied to both mono and bili cohesive density scores. It was found that 42% of mono (5 of 12) and 68% of bili first graders (13 of 19) scored within this range, while 58% of mono (7 of 12) and 32% of bili first graders (6 of 19) scored below. At the same time all three MB first graders achieved a cohesive density within the overlap. Thus, the distribution by L2 preschool experience qualifies the overall distribution in that the overlap between first and fourth graders described earlier is mainly attributable to the bili first graders' cohesive density scores.

From first to fourth grade the mean cohesive density of both mono and bili group increases (Fig. 6.7 and Tab. 6.3). However, the rate of increase varies: The mono group's mean increases comparatively strongly, namely by 0,63 ties per clause, while the bili group's mean density increases only about half as much, namely by 0,34 ties per clause. The MB group's cohesive density score, on the other hand, decreases by 0,24 ties per clause. As a result all three experience groups perform almost the same in fourth grade with respect to their mean cohesive density: Bili fourth graders achieve a mean cohesive density of 3,93 ties per clause, i.e. approximately 39 per 10 clauses, while mono fourth graders achieve a marginally lower mean of 3,87 ties per clause, i.e. also approximately 39 per 10 clauses. That is, the difference between mono and bili group has shrunk to negligible 0,06 ties per clause, i.e. merely 6 cohesive devices in 100 clauses. Nominally, the MB group achieves again the highest mean cohesive density, namely 3,94 ties per clause, but this slight advantage over mono and bili fourth graders is also negligible.

All three groups' interindividual variation becomes even lower from first to fourth grade. The mono group's standard deviation decreases from 0,63 ties per clause (19% of the mean) in first to merely 0,3 (8%) in fourth grade and the bili group's standard deviation from 0,43 (12% of the mean) to 0,34 (9%). The drop in interindividual variation described above is also evident in the range of scores: The bili group's range decreases from 1,78 in first to 1,21 in fourth grade, the mono group's range from 2,25 to 1,07 ties per clause, and the MB group's range from 0,63 to 0,33 ties per clause. This means that – keeping in mind the limited representativeness of the MB group – the differences in interindividual variation are also negligible in fourth grade.

Fig. 6.9 shows the distribution of participants' individual cohesive density scores by experience group (mono and bili only) in fourth grade; the corresponding values are given in Tab. 10.9 and Tab. 10.10 in the Appendix (ch. 10.3).

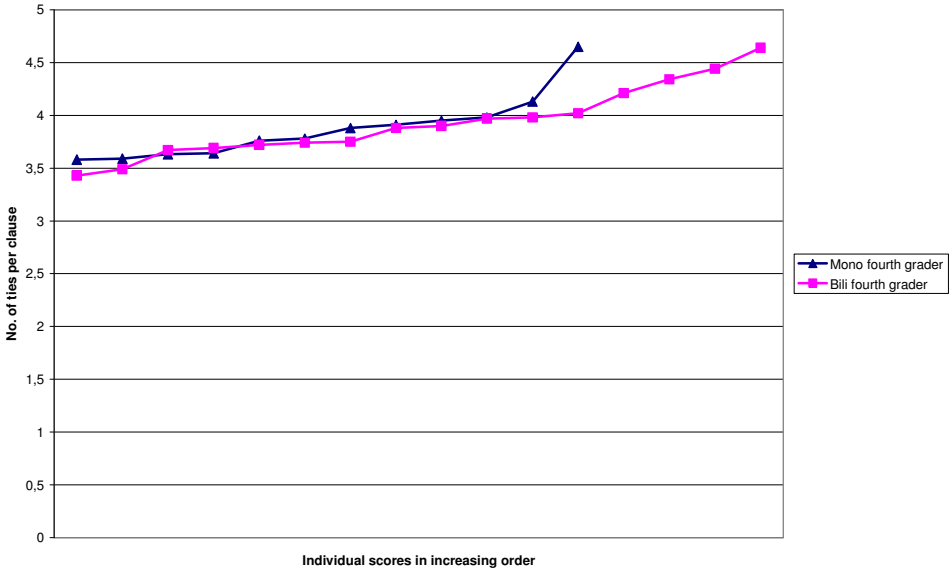

Fig. 6.9 Distribution of participants' individual cohesive density scores by experience group in fourth grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

In fourth grade there is again a strong overlap of cohesive density scores between mono and bili participants. 92% of mono fourth graders (11 of 12) and 88% of bili fourth graders (14 of 16) achieve a cohesive density between 3,58 (mono fourth grade minimum) and 4,64 ties per clause (bili fourth grade maximum). The remaining two bili fourth graders (13%, C5-G4-6, C1-G4-17) realize cohesive density scores below the overlap (3,43 and 3,49, respectively), while the remaining mono fourth grader (8%, C5-G4-12) achieves a cohesive density that is very slightly above the overlap (4,64). Thus, the distribution of cohesive density results in fourth grade shows also virtually no difference between mono and bili participants.

The distribution of mono and bili participants' individual cohesive density scores in first and fourth grade is depicted in Fig. 6.10.

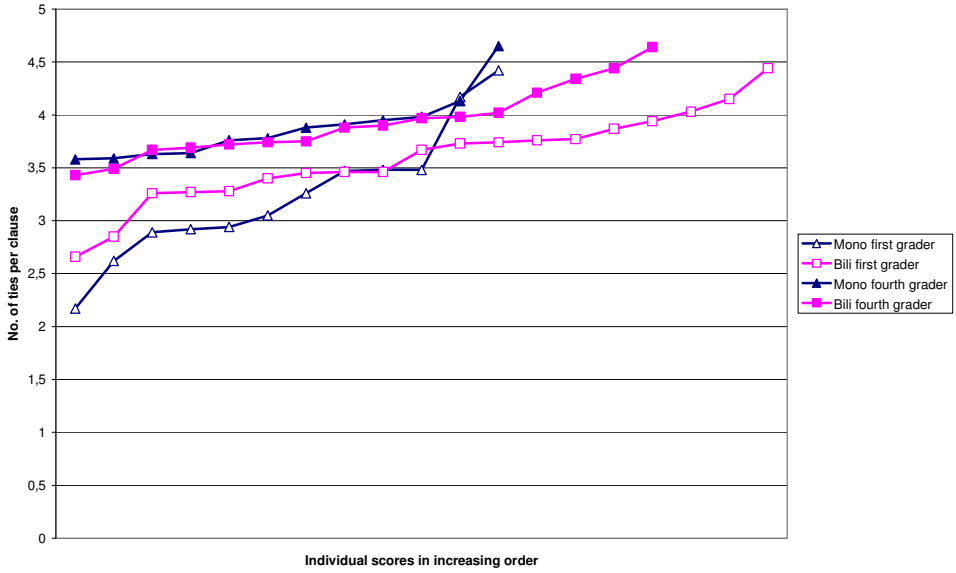

Fig. 6.10 Distribution of participants' individual cohesive density scores by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

Fig. 6.10 shows that the distribution of scores by grade and experience group follows the general trend of a disappearance of the lower first grade scores as opposed to a relative stability in the upper first grade ranges (cf. also the respective minimum scores). However, this development is far more pronounced for the mono than for the bili group, which confirms that the mono group's cohesive density improves more strongly from first to fourth grade. Thus, there is an overlap in mono first and fourth graders' cohesive density between 3,58 (fourth grade minimum) and 4,42 ties per clause (first grade maximum). However, only 17% of mono first graders (C8-G1-16 and C1-G1-8; 2 of 12) achieved a cohesive density within the overlapping range as compared to 92% of fourth graders (11 of 12). The remaining 83% of mono first graders (10 of 12) scored up to 1,41 ties per clause lower, while the one remaining fourth grader (8%, C5-G4-12) achieved a cohesive density slightly above the overlap (4,44). Bili first and fourth grade cohesive density results, on the other hand, overlap in the rage of 3,43 (fourth grade minimum) to 4,44 ties per clause (first grade maximum). 68% of bili first graders (13 of 19) and 94% of fourth graders (15 of 16) achieved a cohesive density within this range. The remaining 32% of bili first graders (6 of 19) achieved a cohesive density below the overlap, while the one remaining fourth grader (6%, C5-G4-18) scored higher (4,64).

To sum up, participants with and without L2 preschool experience performed very similarly in first and fourth grade regarding the cohesive density of their stories

and they followed the general developmental trends outlined earlier. Thus, their mean cohesive density increased from first to fourth grade and the interindividual variation, which was already low in first grade, decreased even further. The distribution of participants' individual cohesive density scores showed a significant overlap between mono and bili participants in both grades, which confirmed this similarity in results. The respective distributions also showed an overlap between first and fourth grade results for both the mono and the bili group. However, the majority of first graders with L2 preschool experience produced as cohesive stories as their fourth grade peers, while this was the case for only a small percentage of mono participants. That is, the development from first to fourth grade was more pronounced for the mono group and the large overlap found between first and fourth graders in the overall results was mainly attributable to the bili group. The small MB group was found to behave somewhat untypically, since its mean cohesive density decreased. This development was largely attributable to their first grade results, however, where they produced a much higher mean density and at the same time had a much lower variation than the other two experience groups.

6.1.4 Statistical results overall cohesive density

General linear model ANOVAs were conducted to statistically test main and interaction effects of grade, sex, and L2 experience group on overall cohesive density (cf. ch. 4.2.3). An analysis including all participants found no main effects of grade ($F(1, 54)=1,83$, ns), sex ($F(1, 54)=1,62$, ns) or experience ($F(2, 54)=1,63$, ns) on the overall results and also no interaction effects among the three factors.

A second ANOVA was conducted without the MB participants, however, since they are the only group whose mean cohesive density decreases from first to fourth grade and this seems to be an atypical behaviour. This second ANOVA showed a significant main effect of grade ($F(1, 51)=7,2$, $p<0,05$), which explained only 12% of the variance in results ($\eta_p^2=0,12$), though. No other main effects were found for sex ($F(1, 51)=0,35$, ns) or experience group ($F(1, 51)=1,5$, ns) and also no interaction effects.

Thus, the statistical analysis found that grade can be a significant factor for explaining variance in cohesive density – if the MB group is excluded. However, even without the MB group the factor grade explains only 12% of the variance in cohesive

density. The statistical analysis also showed that sex and L2 preschool experience are not significant in explaining differences in overall cohesive density.

6.1.5 Summary: Overall cohesive density

This section aimed at answering the question how cohesive participants' stories are (as measured by their degree of cohesive density) and whether there are any differences in cohesion attributable to grade, sex or L2 preschool experience. Based on the findings of previous studies (cf. ch. 3) the following hypotheses had been put forward with respect to these research questions (cf. ch. 2):

(1) Grade/age has a significant influence on the cohesion of participants' stories
(2) Participants' stories become more cohesive from first to fourth grade as measured by the number of cohesive devices
(3) There are no qualitative differences in cohesion between grades as measured by the frequency order of the subcategories of cohesion (lexical ties > references > connectives > ellipses and substitutions).
(4) Sex does not have a significant influence on the cohesion of participants' stories.

The observed results first and foremost showed that all participants, even all first graders, were able to produce at least some cohesive ties. Thus, the lowest number of ties (produced by a first grader) was 2,17, i.e. approximately two ties per clause. At the same time the interindividual differences in cohesive density were found to be very low both in first and fourth grade as compared to those found for coherence.

With respect to an influence of grade the observed results showed two developmental trends, namely

- an increase in cohesive density and
- a (further) decrease in variation.

Thus, participants' mean cohesive density increased from 3,52 in first to 3,91 in fourth grade. That is, first graders produced approximately 7 ties in 2 clauses, while fourth graders produced approximately 4 ties in one or 8 ties in 2 clauses. As opposed to the increase in mean cohesive density participants' standard deviation decreased from 0,56 or 16% of the mean to 0,3 or 8% of the mean. Male and female participants as well as the mono and bili experience groups followed this

development. Only the small MB group performed somewhat untypically – due to extreme results in first grade – in that its mean cohesive density decreased. Correspondingly, the effect of grade became statistically significant only when the MB group was excluded. The overall effect size of grade was low, however, i.e. grade was found to explain only 12% in variance. At the same time the distribution of participants' individual density scores showed a large overlap between first and fourth grade, which helps to explain the low effect size. That is, a large part of first grade stories (62%) were as cohesive as stories produced by fourth graders. The distribution showed that the major change in cohesive density from first to fourth grade concerned the disappearance of the lower first grade scores, indicated by a higher minimum score in fourth grade.

With respect to the influence of sex or experience group (monos vs. bilis) there were no significant observed or statistical differences between groups in either grade regarding mean cohesive density, interindividual variation or the distribution of participants' individual density scores. However, L2 preschool experience had a significant influence on the overlap between first and fourth grade scores. That is, 68% of bili but only 17% of mono first graders produced stories which were as cohesive as those of their fourth grade peers. L2 preschool experience thus also qualified the overlap between first and fourth graders in that the majority of first graders scoring within the fourth grade range (62%, 13 of 21) had L2 preschool experience, while only 24% of them (5 of 21) belonged to the mono group.

To sum up, all four hypotheses were confirmed. The cohesion of participants' stories generally increased from first to fourth grade, while the interindividual variation in cohesion decreased. At the same time a substantial number of first graders produced stories which were as cohesive as those produced by some fourth grader and the main change from first to fourth grade consisted in a reduction of the lower first grade scores. Sex and L2 preschool experience were not found to influence cohesive density except for a stronger overlap of bili first and fourth grade cohesive density as opposed to a more pronounced difference between first and fourth grade cohesion on the part of the monos.

6.2 Cohesive density: Subcategories

The present section aims at answering the question how often participants use the subcategories of cohesion, i.e. references, connectives, substitutions, ellipses and

lexical cohesion, cohesively in first and fourth grade. That is, the degree of the subcategories' density in first and fourth grade is investigated as well as any possible changes from first to fourth grade. Background variables taken into account besides grade are again sex and experience group. In the following, the observable and statistical results as well as a short summary will be given for each subcategory followed by a summary covering all of the subcategories.

6.2.1 References

6.2.1.1 Overall reference density

The present section analyzes how often participants use referential ties[5] in first and fourth grade and whether there are any differences attributable to grade, sex or L2 preschool experience. Tab. 6.4 gives the descriptive statistics for participants' reference density in first and fourth grade.

	Reference density (No. of referential ties per clause)					
	Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	34	1,15	,23	1,17	,57	1,65
Grade 4	32	1,23	,16	1,23	,76	1,62
Total	66	1,19	,20	1,21	,57	1,65

Tab. 6.4 Descriptive statistics reference density

All participants produced at least some referential ties. This is evident from the minimum scores, which show that even first graders produced at least 0,57 referential ties (minimum score), i.e. approximately one reference in two clauses. On average, first graders used 1,15 referential ties per clause, i.e. approximately one reference per clause. First graders' interindividual differences as measured by the standard deviation amount to 0,23 or a variability ratio of 20% in relation to the mean, which is slightly higher than the overall interindividual variation described above. At the same time their results have a range of 1,08 references per clause from minimum (0,57) to maximum score (1,65), i.e. the highest reference density achieved is more than twice the density of the lowest score (Tab. 6.4).

Participants' mean reference density increases only marginally from first to fourth grade. That is, fourth graders produce only eight more references in 100

[5] The terms 'reference' and 'referential tie' will be used synonymously in the following, since only references functioning cohesively were included in the analysis.

clauses (+0,08), namely a mean of 1,23 references per clause. As opposed to participants' mean density their interindividual variation as measured by the standard deviation decreases to 0,16 or 13% of the mean. Similarly, the range of reference density results decreases to 0,86, which is attributable to an increase in the minimum score as opposed to an almost constant maximum.

Fig. 6.11 shows the distribution of participants' individual reference density scores in first and fourth grade in increasing order; the corresponding values are given in Tab. 10.1 and Tab. 10.2 in the Appendix (ch. 10.1).

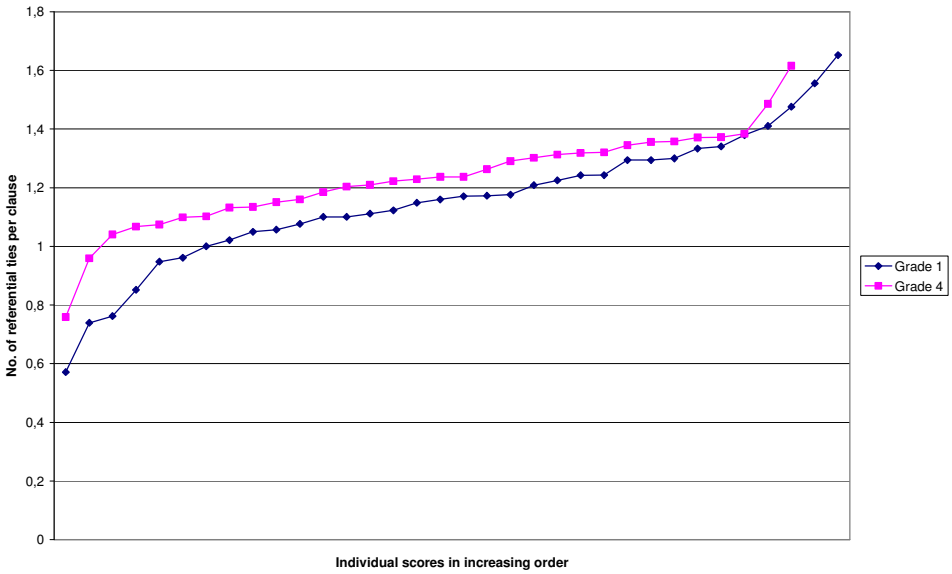

Fig. 6.11 Distribution of participants' individual reference density scores by grade

There is a large overlap in results between first and fourth grade, which ranges from 0,76 (fourth grade minimum) to 1,62 ties per clause (fourth grade maximum). All 32 fourth graders and also 92% of first graders (31 out of 34) obtained a reference density within this range. One first grader (3%, C8-G1-2) obtained a substantially lower reference density (0,57), a second first grader (3%, C1-G1-4) a marginally lower one (0,74) and a third first grader (3%, C8-G1-16[6]) even a marginally higher density than the overlap (1,65). In total the distribution thus seems to indicate that there are no "typical" first and fourth grade scores but that the majority of first and fourth graders perform the same when it comes to reference density.[7]

[6] A female first grader without L2 preschool experience.
[7] Both grades are influenced by somewhat extreme values. Thus, not only first grader C8-G1-2 (3%, 0,57) but also fourth grader C1-G4-17 (3%, 0,76) achieves an extremely low reference density (compared to their respective peers). Even if one or both of these values are removed, however, the number of first graders scoring below the overlap remains small (max. 15% or 5 of 34).

To sum up, participants' mean reference density increases and their interindividual variation decreases from first to fourth grade. However, the mean increases only marginally and at the same time the distribution of participants' individual scores shows close to no differences between first and fourth grade. That is, almost all first graders are likely to produce a reference density that could also have been produced by a fourth grader. Thus, the results show that the difference between participants' first and fourth grade reference density is negligible.

6.2.1.2 Reference density by sex

Tab. 6.5 gives the reference density results obtained by male and female participants in first and fourth grade.

		Reference density (No. of referential ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Male	12	1,24	,20	1,23	,85	1,56
	Female	22	1,11	,23	1,11	,57	1,65
	Total	34	1,15	,23	1,17	,57	1,65
Grade 4	Male	10	1,24	,14	1,30	1,04	1,37
	Female	22	1,22	,17	1,23	,76	1,62
	Total	32	1,23	,16	1,23	,76	1,62

Tab. 6.5 Descriptive statistics reference density by sex and grade

Male first graders achieve a mean density of 1,24 referential ties per clause, while their female counterparts' mean density is slightly lower with 1,11 ties per clause, i.e. female first graders produce roughly 1 reference less in 10 clauses. At the same time male first graders' interindividual variation is a little lower: Their results have a standard deviation of 0,2, which corresponds to a variability ratio of 16%, and females 0,23, i.e. a variability ratio of 21%. This difference in variation is also expressed in the range of scores, which is 0,71 for male and 1,08 for female first graders.

Fig. 6.12 gives the distribution of participants' individual reference density scores for male and female first graders; the corresponding values are given in Tab. 10.3 and Tab. 10.4 in the Appendix (ch. 10.2).

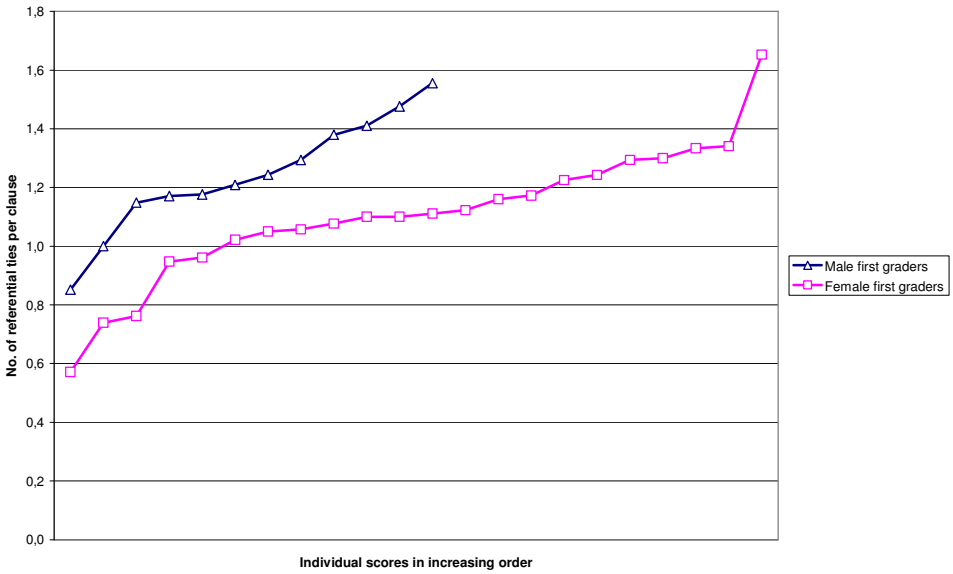

Fig. 6.12 Distribution of participants' individual reference density scores by sex in first grade

Fig. 6.12 shows an overlap of reference density scores between 0,85 (male minimum) and 1,56 references per clause (male maximum). Consequently, all male first graders score within this range. 82% of females (18 of 22) also do so. Three of the remaining female first graders (14%) score below (from 0,57 to 0,76) and one (5%, C8-G1-16) above the overlap (1,65). However, Fig. 6.12 also shows that this last female first grader needs to be considered a statistical outlier due to its extreme value. That is, the overlap between male and female first graders rather ranges from 0,85 to 1,34. Thus, only 67% male first graders (8 of 12) would be considered as scoring within the overlap, while 33% (4 of 12) of them score higher; female results remain as described for the original range. An interpretation in terms of a slight male advantage is made difficult, however, by the large difference in group size. Consquently, the distribution mainly confirms the slight difference in interindividual variation between male and female first graders outlined above. This difference in variation is, however, most likely also attributable to a group size effect.

Only female participants' reference density increases from first to fourth grade, namely by 0,11 referential ties per clause or approximately one tie in 10 clauses. As a result, male and female mean density are even more similar in fourth than in first grade; the difference is now merely 0,02 or 2 lexical ties in 100 clauses. At the same time female fourth grade results are only marginally more heterogeneous than those of their male counterparts. This similarity is indicated by a female standard deviation of 0,17 or 14% of the mean as opposed to 0,14 or 11% of the mean for male fourth

graders. Similarly, females' scoring range from minimum to maximum is 0,86 as opposed to 0,33 for the male fourth graders. Any of these differences in fourth grade are negligible, however, and seem again most likely attributable to a group size effect.

Fig. 6.13 shows the distribution of male and female participants' individual reference density scores in first and fourth grade, i.e. on the one hand the differences and/ or similarities between male and female participants in either grade and on the other hand the differences between each sex's first and fourth grade results. The corresponding numerical values are given in Tab. 10.3 to Tab. 10.6 in the Appendix (ch. 10.2).

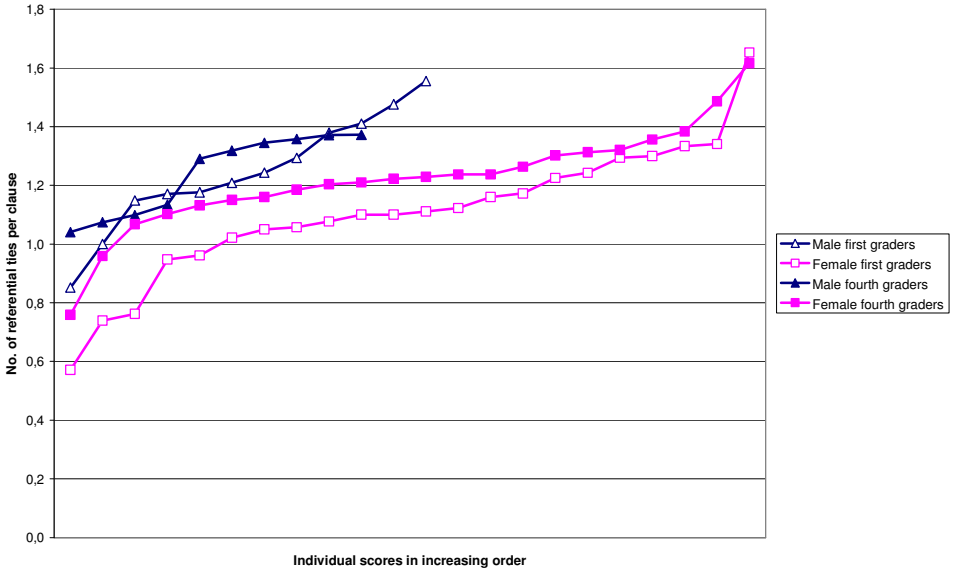

Fig. 6.13 Distribution of participants' individual reference density scores by sex and grade

Male and female results overlap again to a high degree. In fourth grade this overlap ranges from a reference density of 1,04 (male minimum) to 1,37 referential ties per clause (male maximum) and thus includes again all male fourth graders. 77% of female fourth graders (17 of 22) also achieved a reference density in this range, while two female fourth graders (9%) scored below and three of them (14%) up to 0,25 above the overlap. Keeping in mind the respective group sizes this means again that there is little difference between the two groups.

With respect to the distribution of each sex's reference density in first as opposed to fourth grade the following can be observed: Male first and fourth grade results overlap in a range from 1,04 (male fourth grade minimum) to 1,37 referential ties per clause (male fourth grade maximum). All 10 male fourth graders score within

this range but only 50% of first graders (6 of 12). 17% of the remaining male first graders (2 of 12) achieve a reference density below the overlap, while 33% of first graders (4 of 12) realized a cohesive density above the overlap. [8]

Female first and fourth grade results overlap more strongly in a range from 0,76 (fourth grade minimum) to 1,62 (fourth grade maximum), which includes all of the fourth graders (100%) and 86% of first graders (19 of 22). Two female first graders (9%) achieved a (marginally) lower cohesive density (0,57 and 0,74, respectively), while one female first grader (5%, C8-G1-16) achieved a marginally higher score (1,65).[9] These results indicate that male and female participants follow the general trend of largely similar results for reference density in first and fourth grade.

To sum up, no significant observable differences in reference density were found between male and female participants in either grade. Both groups followed the overall trends identified for reference density: Largely similar mean reference densities in first and fourth grade, a strong overlap of first and fourth grade scores, and a decrease from a comparatively low interindividual variation to an even lower one in fourth grade.

6.2.1.3 Reference density by experience group

Tab. 6.6 gives the reference density results obtained in first and fourth grade by participants with bilingual preschool experience (bili), children with monolingual German preschool experience (mono) and participants who went to monolingual German groups in a preschool offering also bilingual groups (MB). As in earlier sections a main comparison will be made between mono and bili results.

[8] However, C1-G1-18 scored 1,38 and C1-G1-11 scored 1,41 references per clause, and thus only marginally higher than the fourth grade maximum (one to four references more in 100 clauses). C8-G1-7 scored 1,48 (+0,11) and C1-G1-2 scored 1,56 references per clause (+0,19), on the other hand, which corresponds to approximately one and two references more in 10 clauses.

[9] As described earlier, however, this latter first grader needs to be considered a statistical outlier. If she is excluded, an overlap in the range of 0,76 to 1,34 referential ties per clause can be identified. This overlap both includes and excludes the same number of first graders as the original one, but means that four of the fourth graders (18%) score (slightly) higher. Similarly, female fourth grader C1-G4-17 seems to produce an exceptionally low reference density compared to her peers (0,76, i.e. the female minimum in fourth grade). If her result is also disregarded, the overlap between first and fourth grade ranges from 0,96 to 1,34 and includes 95% of fourth graders (21 of 22) as well as 77% of first graders (17 of 22). 18% of fourth graders (4 of 22) and 5% of first graders (1 of 22) then score higher, while 18% of first graders (4 of 22) and 5% of fourth graders (1 of 22) score lower than the overlap. The comparison of females' first and fourth grade distribution thus shows only one overall trend, namely that of a substantial overlap.

		Reference density (No. of referential ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Mono	12	1,13	,32	1,20	,57	1,65
	Bili	19	1,15	,15	1,15	,85	1,56
	MB	3	1,28	-)[10]	-	1,05	1,41
	Total	34	1,15	,23	1,17	,57	1,65
Grade 4	Mono	12	1,23	,16	1,23	,96	1,62
	Bili	16	1,20	,17	1,19	,76	1,49
	MB	4	1,35	-	-	1,30	1,37
	Total	32	1,23	,16	1,23	,76	1,62

Tab. 6.6 Descriptive statistics reference density by grade and experience. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

Mono and bili first graders achieve a very similar reference density. That is, bili first graders produce an average of 1,15 referential ties per clause and thus only 0,02 references per clause or 2 more in 100 clauses than the mono group, whose mean reference density is 1,13 per clause. The highest reference density in first grade is obtained by the small MB group with an average of 1,28 references per clause.

The three experience groups differ, on the other hand, in their degree of interindividual variation. Mono first graders have the comparatively largest variation with a standard deviation of 0,32, which corresponds to a variability ratio 28% of the mean, while bili first graders' results have a standard deviation of only 0,15 or of 13% of the mean. Similarly, the mono group has the largest range of scores with a difference of 1,08 ties per clause between the lowest (0,57) and highest reference density of a mono first grader (1,65). The bili group has the second largest range with a difference of 0,71 ties per clause between minimum (0,85) and maximum score (1,56). MB first graders have the comparatively smallest range with only 0,36 ties per clause from minimum (1,05) to maximum referential density (1,41).

The difference between the mono and the bili group's heterogeneity is reflected in Fig. 6.14, which shows the distribution of mono and bili first graders' individual reference density results. The corresponding values are given in Tab. 10.7 and Tab. 10.8 in the Appendix (ch. 10.3).

[10] The MB group is too small for a reliable calculation of standard deviation and median. Thus, only the range of scores will be used as a measure for their interindividual variation.

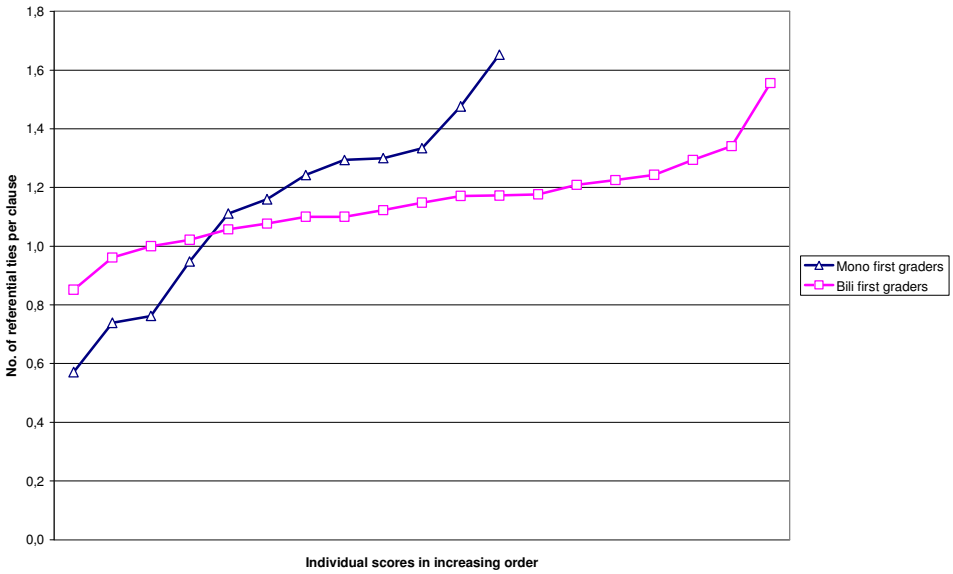

Fig. 6.14 Distribution of participants' individual reference density scores by experience group in first grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience

Mono first graders' larger heterogeneity is evident in the stronger steepness of their curve. At the same time there is a large overlap between mono and bili results, which ranges from 0,85 (bili minimum) to 1,56 (bili maximum). All bili first graders score within this range as compared to 67% of the mono group. 25% of mono first graders (3 of 12) achieve lower reference density scores,[11] while the remaining mono first grader (8%, C8-G1-16) achieves a referential density above the overlap (1,65), which is at the same time the highest referential density of all participants. Thus, the distribution seems to indicate that the mono group has a tendency to score somewhat more often in the lower ranges. Since group sizes are relatively small, though, the percentages correspond to a comparatively low number of participants and this tendency would need to be confirmed in a larger data set.

All three experience groups' mean reference density increases from first to fourth grade, albeit only slightly. The mono group has the comparatively strongest increase with a mean reference density of 1,23 references per clause in fourth grade (+0,1) followed by bili fourth graders, who obtain a mean reference density of 1,2 per clause (+0,05). MB fourth graders obtain a mean of 1,35 (+0,07). As a result of these different increase rates mono and bili fourth graders' reference density is again very

[11] C8-G1-18 produces 0,76 referential ties per clause, i.e. approximately one in 10 clauses less, C1-G1-4 scores 0,74, i.e. also roughly one in 10 clauses less, and C8-G1-2 produces 0,57, i.e. 0,28 or approximately three references less in 10 clauses.

similar – they differ by only 0,03 or 3 references in 100 clauses. The small group of MB fourth graders scores a little higher than mono and bili group.

The interindividual variation of mono and MB experience group (as measured by standard deviation and range) decreases from first to fourth grade (Tab. 6.6). Thus, mono fourth graders' standard deviation is 0,16, which corresponds to 13% of the mean – less than half as much as in first grade – and their range of results decreases to a difference of 0,66 between minimum and maximum score. This difference in range is largely due to a higher minimum as opposed to a similar (even slightly lower) maximum score. The MB group's range also drops, namely to 0,07 references per clause. The bili group's standard deviation (and correspondingly the range), on the other hand, rises slightly from 0,15 in first grade to 0,17 references per clause in fourth grade. When compared in relation to the mean, however, fourth grade (14%) and first grade variation (13%) are only marginally different so that it can be said that the interindividual variation of both mono and MB group decreases, while the one of the bili group remains constant. Thus, MB fourth grade results are the most homogeneous followed by mono and bili fourth graders, which have roughly the same variation (13% and 14% of the mean, respectively).

Fig. 6.15 shows the distribution of mono and bili participants' individual density scores as a function of grade and experience group, i.e. on the one hand the differences and/or similarities between mono and bili participants in first and fourth grade and on the other hand the difference between either experience group's first and fourth grade results. The corresponding values are given in Tab. 10.7 to Tab. 10.10 in the Appendix (ch. 10.3).

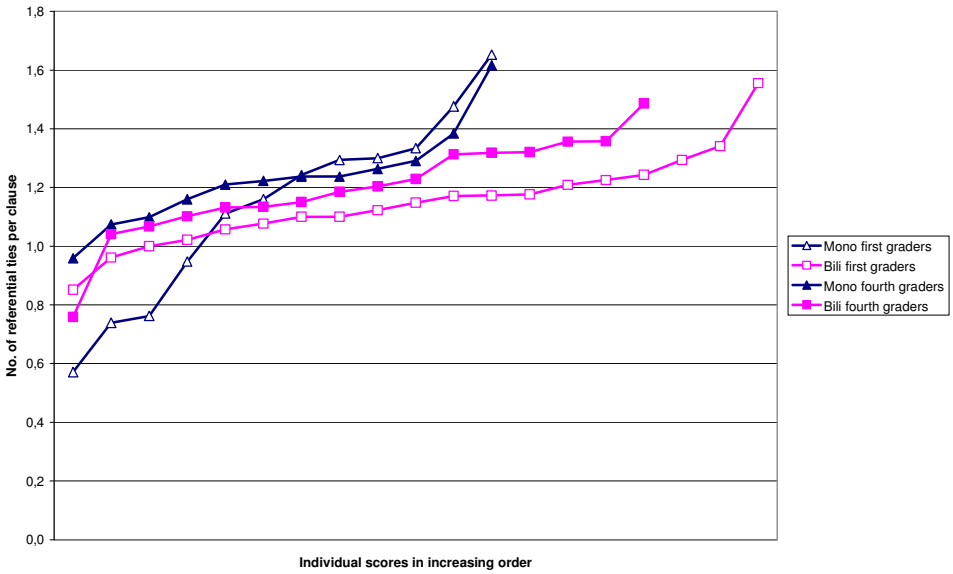

Fig. 6.15 Distribution of participants' individual reference density scores by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience

A large overlap in both groups' fourth grade results is evident; it ranges from 0,96 (mono minimum) to 1,49 references per clause (bili maximum). 92% of mono fourth graders (11 of 12) as well as 94% of bili fourth graders (15 of 16) achieve a reference density within this range. The remaining mono fourth grader (8%, C5-G4-12) obtains a somewhat higher referential density (1,62), while the remaining bili fourth grader (6%, C1-G4-17) achieves a density below the overlap (0,76). Thus, the distribution confirms mono and bilis' overall similarity of results and interindividual variation in fourth grade.

As to each experience group's first and fourth grade results, the following can be observed: There is a large overlap of mono first and fourth grade scores, which ranges from 0,96 (mono fourth grade minimum) to 1,62 (mono fourth grade maximum). All mono fourth graders obtained a reference density within this range as opposed to only 58% of first graders (7 of 12). 33% of the remaining mono first graders (4 of 12) scored lower (one of them marginally), while one first grader (8%) scored higher (1,65; the already mentioned outlier C8-G1-16). That is, the main difference between the mono group's first and fourth grade results is an increase in the lower first grade scores (minimum density) as it was already observed for participants' overall cohesive density.

An even larger overlap between first and fourth grade can be observed for the bili group. This overlap ranges from 0,85 (bili first grade minimum) to 1,49 referential ties per clause (bili fourth grade maximum). 95% of bili first graders (18 of 19) and

94% of bili fourth graders (15 of 16) scored within this range. One first grader (5%, C1-G1-2) produced more (1,56) and one fourth grader (6%, C1-G4-17) fewer referential ties than the range of the overlap (0,76). However, this latter bili fourth grader seems to produce an exceptionally low reference density compared to her peers. If her result is excluded, bili first and fourth grade scores overlap from 1,04 to 1,49; all bili fourth graders are included in this range, while 21% of first graders score below the overlap. Thus, a comparison of the bili group's first and fourth grade distribution does not initially reveal any difference between the two grades. However, this changes once an outlier is disregarded. Bili first and fourth grade curves then point to a similar trend as found for the monos, namely an increase of the minimum score as the main difference between first and fourth grade distribution (paired with a strong overlap).

To sum up, the difference in mean reference density between children with and without L2 preschool experience was found to be negligible both in first and fourth grade, while the MB group achieved a slightly higher mean than the other two groups in both grades. With respect to interindividual variation in first grade, however, the mono group was found to be most heterogeneous followed by the bili and then the small MB group. From first to fourth grade all three groups' mean reference density increased. At the same time the interindividual differences of mono and MB group decreased, while the variation of the bili group remained fairly stable. As a result there was also almost no difference in interindividual variation between experience groups in fourth grade. The only effect attributable to L2 preschool experience was found in the distribution of participants' individual results, which showed first of all a greater likelihood of mono first graders to produce low reference density scores. The distribution also showed a more pronounced difference between the mono group's first and fourth grade curves, i.e. a stronger increase of the mono group's minimum reference density. The bili curves, on the other hand, showed a similar difference between first and fourth grade only after a fourth grade outlier was removed from the analysis.

6.2.1.4 Statistical results referential density

A general model ANOVA including all participants found no significant main effects attributable to grade ($F(1, 54)=0,12$, ns) or experience ($F(2, 54)=0,91$, ns) as well as no interaction effects. The main effect for sex missed significance ($F(1, 54)=3,1$,

p=0,09, ns) but it indicated a trend toward significance. Since male participants are the only group investigated whose mean reference density does not increase from first to fourth grade, a subsequent independent samples t-test was conducted to further explore this trend. The t-test showed that the overall trend is attributable to the two sexes' first grade results, where the difference between male (M=1,24) and female mean reference density (M=1,11) barely missed significance (t(32)=2, p=0,053, ns). In fourth grade, on the other hand, not even a trend was found (t(30)=0,43, ns). Since all other groups seemed to behave consistently with their mean reference density increasing from grade 1 to 4 (see above), no additional analyses were conducted.

Thus, the statistical analysis showed that neither grade nor L2 preschool experience are statistically significant factors for variation in referential density results. Similarly, sex cannot be considered a significant influence factor, even if some trends towards significance were observed; these trends would, however, have to be confirmed in a larger data set.

6.2.1.5 Summary: Referential density

The present section aimed at answering the question how often participants use referential ties and whether there are any differences attributable to grade, sex or L2 preschool experience. All participants used at least some references – even first graders produced a minimum cohesive density of roughly one reference in two clauses. At the same time participants' interindividual variation was slightly higher than the one found for overall cohesive density but still low compared to, for example, the coherence results – even if all subgroups' results were influenced by statistical outliers.

No significant observed or statistical differences were found between first and fourth grade with respect to participants' mean reference density; it increased only marginally from first to fourth grade. Thus, first graders produced an average of 1,15 and fourth graders 1,23 referential ties per clause – a difference of merely 0,08 or 8 references in 100 clauses. A comparison of participants' distribution curves reflected the small difference between grades, since it showed a large overlap between first and fourth grade scores: Almost all first graders (92% with and 85% without outliers) achieved a lexical density which could also have been produced by a fourth grader. Consequently, no "typical" first or fourth grade reference density could be identified.

While the mean density remained almost the same, participants' interindividual variation, which was already comparatively low in first grade, decreased even further until the end of fourth grade (from a standard deviation of 0,23 or 20% of the mean to 0,16 or 13% of the mean). The distribution of participants' scores also reflected the decrease in variation, since it showed that the main difference between participants' first and fourth grade distribution curves consisted in a higher minimum in fourth grade.

Neither sex nor L2 preschool experience was found to have any significant observable or statistical influence on participants' mean reference density. That is, male and female mean density differed only slightly in first grade and was virtually the same in fourth grade. Mono and bili mean density was also virtually the same in both grades; only the small MB group achieved a slightly higher density than the other two. With respect to participants' interindividual differences male and female participants performed very similarly in both grades, even if females' variation was slightly higher in first grade (mainly due to two extreme values). Some differences in variation were observed between mono and bili participants, however. Thus, mono first graders' results differed somewhat more than those of their bili peers (the mono group's standard deviation was 0,32, i.e. a variability ratio of 28%, as opposed to the bili's standard deviation of 0,15, i.e. a variability ratio of 13%). Almost all subgroups adhered to the general developmental trends identified for reference density, namely an only marginal increase in mean density and a decrease in variation. The only exceptions were male participants' mean density and the bili group's interindividual variation, which both remained stable.

The distribution of participants' individual scores showed also no differences between males and females in first or fourth grade or in regard to either groups' difference between first and fourth grade. Both sexes followed the general trend of a strong overlap between first and fourth grade scores and an increase in minimum score as the major change from first to fourth grade. L2 preschool experience, on the other hand, seemed to have some effect on the distribution of individual scores. Thus, 25% of mono first graders (3 of 12) performed below a common overlap, which means that mono first graders were somewhat more likely than bili first graders to produce low referential density scores. However, due to the small absolute numbers this finding would need to be corroborated in a larger data set. By the end of fourth grade this difference between monos and bilis had disappeared. However, mono first

and fourth grade distributions were also observed to differ more than those of bili participants, which is again attributable to mono first graders' greater likelihood to produce lower scores. One could also say that mono participants' minimal reference density increased more strongly.

To sum up, neither grade, nor sex, nor experience were found to have a significant impact on reference density and its development. Instead, first and fourth graders are virtually indistinguishable with respect to their use of referential ties and the main difference between first and fourth grade was a decrease in interindividual variation (marked by a higher minimum score in fourth grade). Male and female participants performed almost the same in either grade and in their development from first to fourth grade. L2 preschool experience was merely found to be responsible for a somewhat lower interindividual variation in first grade; otherwise the three experience groups also performed largely the same.

6.2.2 Connectives

6.2.2.1 Overall connective density

The present section aims at answering the question how often participants use connective ties[12] and whether there are any differences attributable to grade, sex or L2 preschool experience. Tab. 6.7 gives the descriptive statistics for participants' connective density in first and fourth grade.

	Connective density (No. of connective ties per clause)					
	Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	34	,48	,29	,49	0	,97
Grade 4	32	,62	,12	,58	,46	,91
Total	66	,55	,23	,56	0	,97

Tab. 6.7 Descriptive statistics connective density by grade

As the minimum score shows not all first graders produce connective ties, even if they achieve a mean density of 0,48 connectives per clause. That is, on average first graders connect almost every two clauses with a connective. At the same time the variation in connective density is quite large. This is reflected, on the one hand, by the range of connective density from minimum (0) to maximum score (0,97), which

[12] The terms 'connective' and 'connective tie' will be used synonymously in the following, since only connectives functioning cohesively were included in the analysis.

indicates that some first graders produce almost one connective per clause while others do not realize any connectives. Additionally, the standard deviation is very large with 0,29 connectives per clause or a variability ratio of 60%.

From first to fourth grade participants' mean connective density increases to 0,62 connective ties per clause, i.e. fourth graders realize 0,14 connective more per clause or roughly one connective per 10 clauses. Thus, not only do all fourth graders produce connectives (minimum score 0,46) but on average they connect at least every two clauses with a connective. While the mean connective density increases, the interindividual differences drop sharply from first to fourth grade: Fourth graders' results have a standard deviation of 0,12 or 19% of the mean.[13] Similarly, the range of connective density results decreases from 0,97 to 0,45, i.e. to less than half the first grade range. As first and fourth grade minimum indicate this decrease in range seems to be due to an increase in the minimum score (from 0 to 0,46) as opposed to a fairly constant maximum, which is even slightly lower in fourth grade (0,97 vs. 0,91).

Fig. 6.16 shows the distribution of individual connective density results in first and fourth grade; the corresponding values are given in Tab. 10.1 and Tab. 10.2 in the Appendix (ch. 10.1).

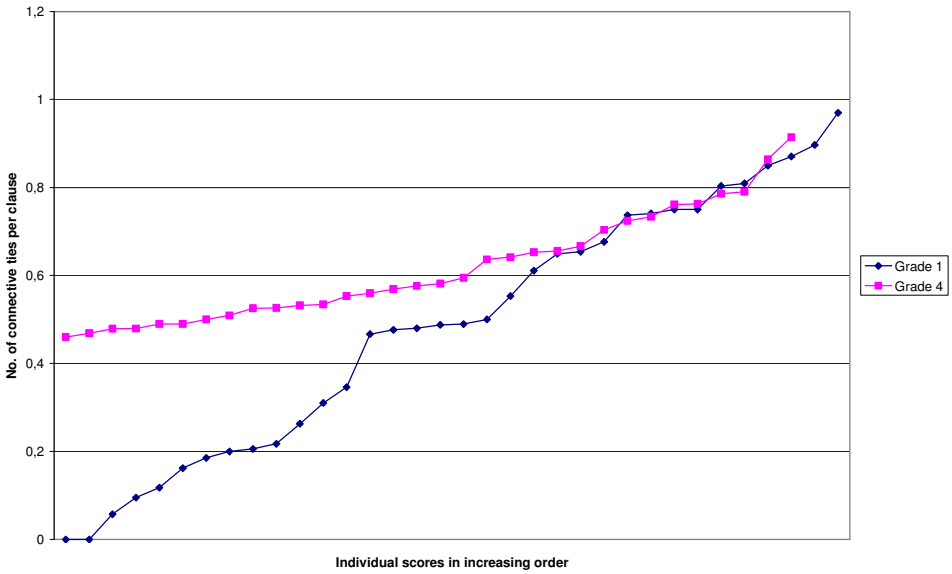

Fig. 6.16 Distribution of participants' individual connective density scores by grade

The distribution of participants' individual connective density scores shows first of all that two of 34 first graders (6%, C8-G1-8 and C8-G1-16) do not produce any

[13] However, this variability ratio is still high compared to the fourth grade variation found for overall cohesive density (8%) and it is also a little higher than the fourth grade variation in reference density (13%).

connective ties, while the great majority of them (94%) produces at least some such ties – even if the steepness of the curve simultaneously points to the large interindividual differences. The distribution also reflects that all fourth graders produce at least some connectives and that their minimum lies much higher (cf. the fourth grade minimum score). At the same time there is an overlap of first and fourth grade results in the range of 0,46 (fourth grade minimum) to 0,91 connectives per clause (fourth grade maximum). All fourth graders achieve a connective density within this range as opposed to 60% of first graders (20 of 34). 38% of the remaining first graders producing connectives (13 of 34) obtained a connective density below that range.[14] One first grader (3%, C1-G1-8) produced a higher number of connective ties per clause, namely 0,97, which is additionally the largest score of all participants. Thus, the distribution confirms the assumption made above that the main difference between first and fourth grade is the increase of the lower scoring ranges as opposed to fairly constant results in the upper ones.

To sum up, all fourth but not all first graders produce connectives and the mean number of connectives increases from first to fourth grade, while the variation decreases very strongly. The distribution of participants' individual connective density scores shows a substantial overlap between first and fourth grade results. Together with the minimum and maximum scores it also shows that the main difference between the two grades involves an increase in the lower scoring ranges as opposed to a relative stability in the upper ones – just as found for participants' overall cohesive density.

6.2.2.2 Connective density by sex

Tab. 6.8 gives the connective density results obtained by male and female participants in first and fourth grade.

[14] Between 0,11 and 0,46 connectives less per clause.

		Connective density (No. of connective ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Male	12	,57	,28	,64	,10	,90
	Female	22	,44	,29	,47	0	,97
	Total	34	,48	,29	,49	0	,97
Grade 4	Male	10	,62	,11	,59	,49	,79
	Female	22	,62	,13	,58	,46	,91
	Total	32	,62	,12	,58	,46	,91

Tab. 6.8 Descriptive statistics connective density by sex and grade

All male first graders realized at least some connective ties, i.e. a minimum of 0,1 per clause or 1 tie every 10 clauses. Not all female first graders, however, realized connective ties; this is evident from their minimum score of 0. Male first graders achieve a mean density of 0,57 connectives per clause, while female first graders obtain a slightly lower connective density of 0,44 connectives per clause. Thus, male first graders realize 0,13 connective ties more per clause or approximately one per 10 clauses – even if both sexes produce approximately one connective tie per clause.

At the same time both groups' interindividual variation is very large: Male first graders' standard deviation amounts to 0,28 or 49% of the mean. Female first graders' results vary even more with a standard deviation of 0,29 or 66% of the mean. Similarly, both groups' range between minimum and maximum score is very large. Male first graders' results have a range of 0,8 between their group members' minimum (0,1) and maximum connective density (0,9); female first graders results differ again even more with a range of 0,97 between minimum (0) and maximum connective density (0,97).

Fig. 6.17 shows the distribution of male and female participants' individual connective density scores in first grade; the corresponding values are given in Tab. 10.3 and Tab. 10.4 in the Appendix (ch. 10.2).

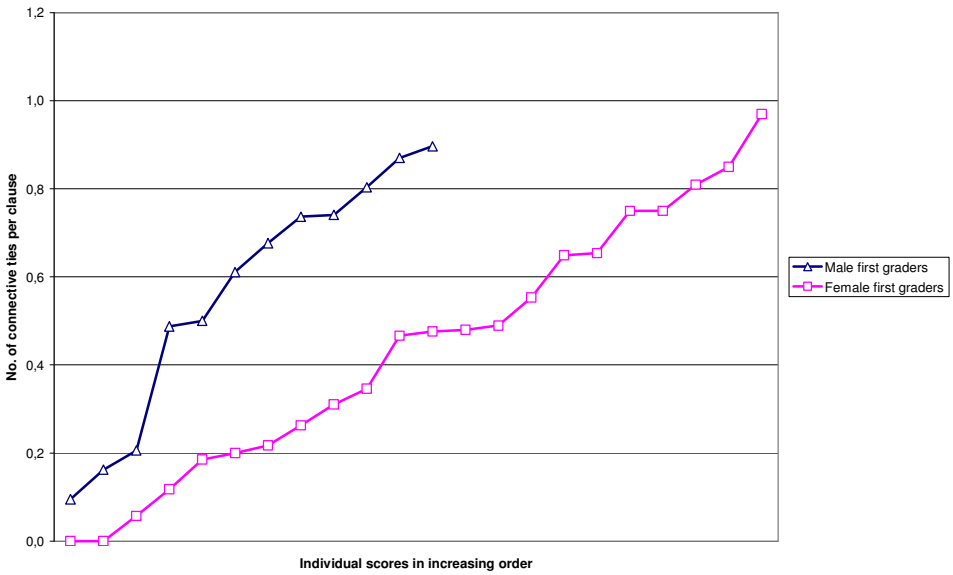

Fig. 6.17 Distribution of participants' individual connective density scores by sex in first grade

The comparatively strong steepness of the distribution curves reflects first of all both sexes' strong interindividual variation. At the same time a large overlap in results is evident, which ranges from 0,1 (male minimum) to 0,9 (male maximum). The overlap encompasses all 12 male and the majority of female first graders (82%, 18 of 22). However, three female first graders (14%) achieved a connective density below the overlap and, as Fig. 6.17 shows, two of those (9%) produced no connectives at all. The remaining female first grader obtained a connective density above the overlap. Keeping in mind the distinct group sizes the distribution thus shows no significant differences between male and female first graders with respect to connective density.

Both male and female participants' mean connective density increases from first to fourth grade. While the male group's density increases by only 0,05 connectives per clause, i.e. 5 more connectives in 100 clauses, females' mean density increases by 0,18 or approximately 2 connectives in 10 clauses. As a result male and female fourth graders achieve the same mean connective density of 0,62 ties per clause. That is, both male and female fourth graders use a connective slightly more often than every two clauses and they realize an average of approximately 6 connectives in 10 clauses. At the same time all male as well as all female fourth graders produce at least some connectives; this is evident again from the minimum score, which is now above zero not only for the male (0,49) but also for the female group (0,48).

While male and female mean density increase, the heterogeneity in their results decreases. Female results' standard deviation drops from 0,29 to 0,13, i.e. from a

variability ratio of 66% to 21%, while male participants' standard deviation decreases from 0,28 to 0,11, i.e. from a variability ratio of 49% to 18% of the mean. Similarly, the range of results from minimum to maximum score drops from 0,97 to 0,45 for the female and from 0,8 to 0,3 for the male group. Consequently, the interindividual differences in fourth grade are very similar for both males and females.

Fig. 6.18 shows the distribution of male and female participants' individual connective density scores in first and fourth grade, i.e. on the one hand the differences and/ or similarities between male and female participants in grade 1 and 4 and on the other hand the differences between the two sexes' first and fourth grade results. The corresponding numerical values are given in Tab. 10.3 to Tab. 10.6 in the Appendix (ch. 10.2).

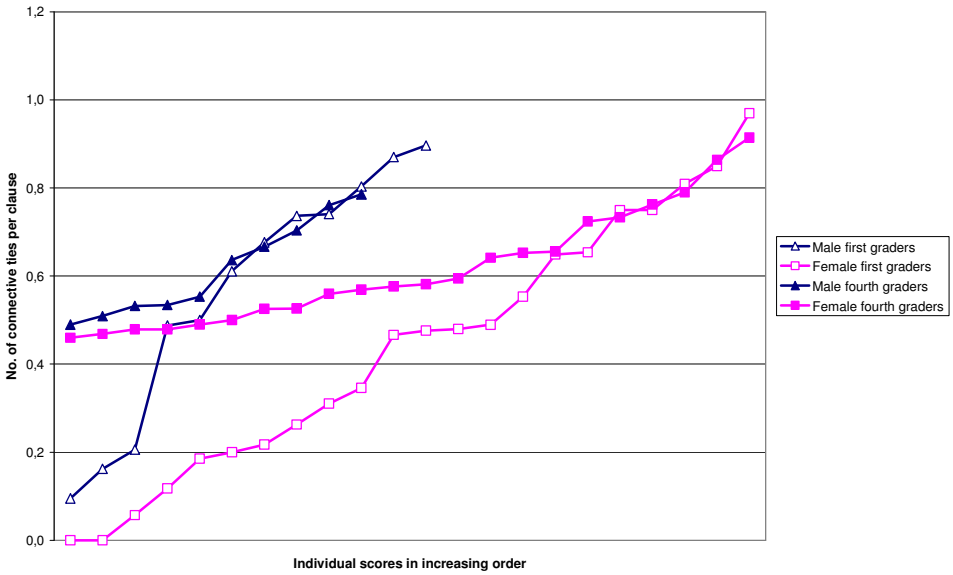

Fig. 6.18 Distribution of participants' individual connective density scores by sex and grade

Male and female fourth grade curves indicate again a large overlap in connective density results. This overlap ranges from 0,49 (male minimum) to 0,79 (male maximum) and consequently includes all male fourth graders. Less but still the majority of female fourth graders (73%, i.e. 16 of 22) achieve a connective density within this range. 19% of female fourth graders (4 of 22) score only negligibly lower (up to 0,03 connectives per clause), while 9% of them (2 of 22) achieve a connective density which lies a little above the overlapping range (0,86 and 0,91, respectively). All in all, however, these differences in fourth grade are negligible, especially when keeping in mind the large difference in group size.

As to either sex's difference between first and fourth grade, the following can be observed: The steepness of each sex's curve lessens from first to fourth grade, which reflects the decrease in interindividual variation described earlier. The curves also reflect that both sexes' fourth grade minimum scores are much higher than the ones in first grade (cf. also Tab. 6.8). Nevertheless, an overlap between first and fourth grade scores can be observed for either sex. Thus, female first and fourth grade connective density results overlap in the range of 0,46 (fourth grade minimum) to 0,91 connectives per clause (fourth grade maximum). All female fourth graders score within this range but also half of the first graders (50%, 11 of 22). Almost half of the female first graders (45%, 10 of 22), however, obtained a lower connective density. One female first grader (5%, C1-G1-8) produced slightly more connectives per clause (0,97). Male first and fourth grade results, on the other hand, overlap between 0,49 (fourth grade minimum) to 0,79 (fourth grade maximum), which also includes all of the fourth graders but only half of the first graders' results (50%, i.e. 6 of 12). Three of the remaining male first graders (25%) produced substantially fewer connectives per clause than the overlap,[15] while three others (25%) produced (slightly) more.[16] All in all this comparison between each sex's first and fourth grade results indicates that fourth graders do not necessarily achieve a higher connective density but that the general trend of an increase in the lower scoring ranges paired with relatively little change in the (first grade) upper scoring ranges applies to both male and female results.

To sum up, male and female participants perform quite similarly in both grades, especially when keeping in mind a likely group size effect. In first grade slight differences can be observed in that female first graders produce a little fewer connective ties per clause and their results vary more. Due to distinct increase rates with regard to mean density and distinct decrease rates with regard to interindividual variation, however, there are close to no differences between the two sexes in fourth grade. The respective distributions not only confirm these findings but also that the general development observed for connective density, namely an elimination of the lower first grade scores paired with a substantial overlap of first and fourth grade scores, holds true for both sexes' results.

[15] These three male first graders produced between 0,28 and 0,39 less connectives than the overlap.
[16] These three other male first graders produced between 0,01 and 0,11 more connectives than the overlap.

6.2.2.3 Connective density by experience group

Participants' connective density results by experience group and grade are given in Tab. 6.9. As in earlier sections, a first main comparison will be made between participants with and without L2 preschool experience (mono vs. bili group), while the small group of children from a monolingual group in a mixed preschool (MB group) will be considered separately.

		Connective density (No. of connective ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Mono	12	,32	,31	,21	0	,97
	Bili	19	,53	,23	,55	,06	,87
	MB	3	,85	-	-	,80	,90
	Total	34	,48	,29	,49	0	,97
Grade 4	Mono	12	,62	,13	,58	,46	,86
	Bili	16	,63	,13	,64	,47	,91
	MB	4	,52	-	-	,48	,55
	Total	32	,62	,12	,58	,46	,91

Tab. 6.9 Descriptive statistics connective density by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

Bili first graders achieve a mean connective density of 0,53 and thus realize 0,21 connectives per clause (or 2 in 10 clauses) more than mono first graders, who reach a mean connective density of 0,32. That is, bili first graders produce approximately one connective every two clauses but mono first graders one connective every three clauses. At the same time all bili but not all mono first graders use at least some connective ties; this is evident from the two groups' minimum scores (0,06 vs. 0).

Mono first graders also show an extremely high variation in results with a standard deviation of 0,31, which corresponds to a variability ratio of 97%, and a range of 0,97 connectives per clause from minimum (0) to maximum score (0,97). Bili first graders' results are also very heterogeneous but still less so than those of the monos: The standard variation for bili first graders is 0,23 or 43% of the mean and the range of scores is 0,81 from minimum (0,06) to maximum (0,87).

Fig. 6.19 shows the distribution of mono and bili participants' individual connective density scores as a function of experience group in first grade; the corresponding numerical values are given in Tab. 10.7 and Tab. 10.8 in the Appendix (ch. 10.3).

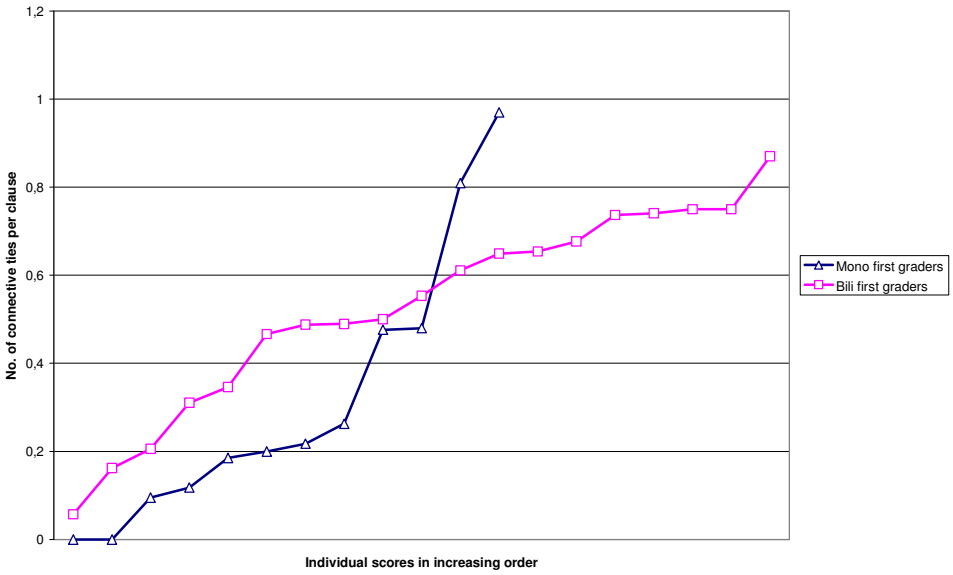

Fig. 6.19 Distribution of participants' individual connective density scores by experience group in first grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience

The steepness of both groups' curves first of all reflects their large interindividual variations. The curves also show an overlap of results from 0,1 (mono minimum) to 0,87 connectives per clause (bili maximum). 95% of bili first graders (18 of 19) achieve a cohesive density within this range as compared to 75% of mono first graders (9 of 12). The remaining bili first grader (5%, 1 of 19) scores slightly lower (0,06). Two mono first graders (17%, 2 of 12) did not produce any connectives at all,[17] while one mono first grader (8%, C1-G1-8) achieved a connective density above the overlap (0,97). However, the two highest mono scores (0,81 and 0,97) seem to be rather exceptional for this group, since the next lowest score amounts to about half of this connective density (0,48) so that scores up to 0,48 would seem more typical for the mono group.[18] Consequently, the overlapping range between mono and bili first graders would rather need to be defined as 0,1 (mono minimum) to 0,48 (highest mono score excluding the two outliers). This new overlap would then include all eight mono first graders who produced connectives but only 32% of bili first graders (6 of 19); 68% of the bilis (13 of 19) would score above the overlap. The first

[17] C8-G1-8 and C8-G1-16 produced 0 connectives in 12 and 23 clauses, respectively.
[18] This skewed distribution of the mono groups' connective density scores is also evident from their median (0,21) as opposed to the mean (0,32) (cf. Tab. 6.9).

grade distribution of connective density scores without mono outliers would thus need to be interpreted as indicating an advantage of the bili group.[19]

MB first graders (N=3) outperform both other experience groups in realizing an average of 0,85 connectives per clause, i.e. they use a connective in almost every clause or roughly 4 connectives in 5 clauses. At the same time the MB group's results are the most homogeneous with a range of only 0,1 between minimum (0,8) and maximum score (0,9). The MB group's minimum score is also the highest one of all three experience groups, while their maximum score (0,9) is close to the other two groups. Keeping in mind its extremely small size, however, the MB group's high performance and low variation cannot be considered representative.

The mono and the bili groups' mean connective density increases to different degrees from first to fourth grade. The mono group's mean connective density rises by 0,3 connectives per clause and thus more strongly than the bili group's mean, which increases by only 0,1. As a result mono and bili fourth graders perform almost indistinguishably with a mean connective density of 0,62 and 0,63, respectively. That is, both groups produce approximately 3 connective ties in 5 clauses; one could also say they connect every two clauses with a connective and produce roughly one additional connective tie in 10 clauses. At the same time not only all bili but also all mono fourth graders use at least some connective ties; this is evident from a rise in the mono minimum score from 0 in first to 0,46 in fourth grade. As mono and bili means increase the MB group's mean decreases by 0,33 connectives per clause and thus MB fourth graders produce fewer connectives than the other two groups, namely 0,54 per clause.

All three experience groups' results become more homogeneous from first to fourth grade. This increase in homogeneity is especially evident for monos and bilis, whose results had an extremely high variation in first grade: The mono group's standard deviation drops from 0,31 or 97% of the mean in first to 0,13 or 21% of the mean in fourth grade. Relatedly, their range of scores decreases from 0,97 to 0,4. The bili group's standard deviation drops from 0,23 or 43% of the mean in first to equally 0,13 or 21% of the mean in fourth grade and their scoring range decreases from 0,81 to 0,44. This means that the interindividual variation in fourth grade is the same for both monos and bilis. The MB group's scoring range decreases only

[19] However, such an advantage would need to be confirmed by a larger number of participants, since only 10 mono but still 19 bili first graders remain and thus the possibility of a group size effect cannot be excluded.

marginally from 0,1 to 0,07 so that it can be said that there is no difference in the MB's interindividual variation in first and fourth grade and they remain the most homogeneous group in this sense.

Fig. 6.20 shows the distribution of mono and bili participants' individual connective density scores in increasing order as a function of grade, i.e. on the one hand the differences and/ or similarities between mono and bili participants in grade 1 and 4 and on the other hand the differences between the two experience groups' first and fourth grade results. The corresponding values are given in Tab. 10.7 to Tab. 10.10 in the Appendix (ch. 10.3).

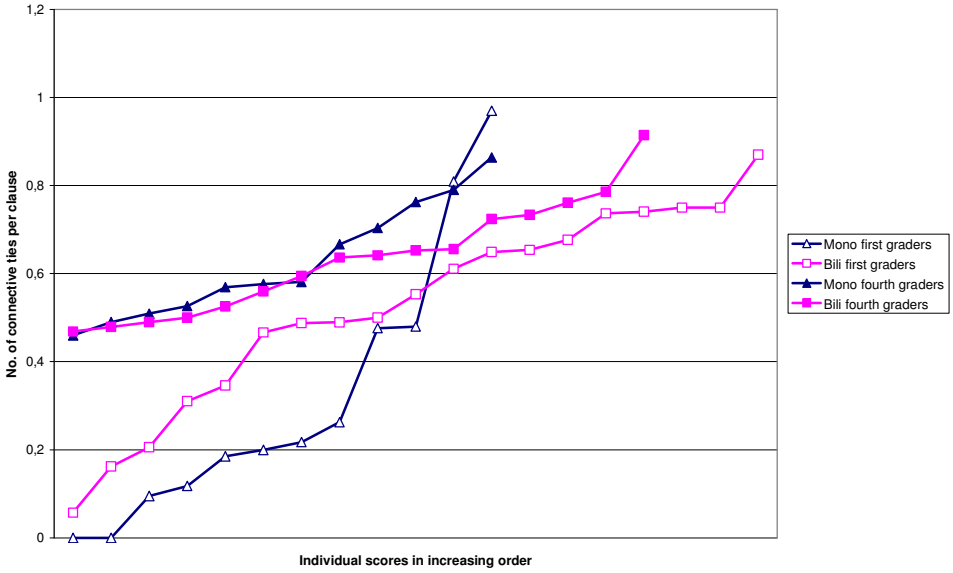

Fig. 6.20 Distribution of participants' individual connective density scores by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

The fourth grade curves show an overlap between mono and bili fourth grade scores, which ranges from 0,47 (bili minimum) to 0,86 (mono maximum). It includes 92% of mono (11 of 12) and 94% of bili fourth graders (15 of 16), i.e. the large majority of both experience groups. Only one mono fourth grader (8%, C5-G4-18) achieved a marginally lower connective density (0,46) and one bili fourth grader (6%, C5-G4-18) a marginally higher connective density (0,91). That is, the fourth grade distribution of participants' connective density results shows no difference between monos and bilis.

A comparison of both experience groups' first and fourth grade curves first of all reflects the strong decrease in interindividual variation described above, since the fourth grade curves are much less steep than their first grade counterparts. The curves also indicate that the decrease in variability is mainly due to a strong increase

in minimum scores (cf. also Tab. 6.9). At the same time the distribution shows an overlap between first and fourth grade results for both groups. Thus, mono results overlap between 0,46 (fourth grade minimum) and 0,86 (fourth grade maximum). All mono fourth graders score within this range, while only 25% of the first graders (3 of 12) do so. 67% of mono first graders (8 of 12) achieve a lower connective density; the one remaining first grader (8%) scores higher than the overlap. Bili first and fourth grade results overlap in the very similar range of 0,47 (fourth grade minimum) to 0,87 (first grade maximum). Almost all bili fourth graders (94%, 15 of 16) achieve a connective density within that range; one fourth grader (6%, C1-G1-8) obtains a higher density. The majority of bili first graders, namely 74% (14 of 19), also scores within this range; 26% of them (5 of 19) obtain a lower connective density. This means that both mono and bili results follow the general trend of an overlap between first and fourth grade and a decrease of results in the lower scoring ranges with again little change in the upper ranges. For the mono group this trend is more pronounced, however, due to larger differences between first and fourth grade results.

To sum up, all bili but not all mono first graders produced connectives. Additionally, bili first graders produced a higher mean connective density and their variation in results was much lower than that of mono first graders. The distribution also pointed towards an advantage of the bili group in that most bili first graders scored above a common overlap once two extreme values of the mono group were disregarded. The small group of MB first graders outperformed both other experience groups with respect to mean connective density and variation. Mono and bili participants followed the general trend of an increase in mean connective density and a (strong) decrease in interindividual variation, even if the rate of increase/ decrease was stronger for the mono group. The small MB group's mean, on the other hand, decreased and their variation remained largely stable. In fourth grade mono and bili participants performed very similarly in regard to mean connective density and variation – this result was also confirmed by the distribution curves – while the MB group had a slightly lower mean density and still the lowest variation of all three groups. Mono and bili groups' first and fourth grade distributions showed that they follow the general trend of an overlap of first and fourth grade results and an increase mainly in the lower density ranges. However, the difference between first and fourth grade was found to be more pronounced for the mono group.

6.2.2.4 Statistical results connective density

General linear model ANOVAs were conducted to statistically test main and interaction effects of grade, sex, and L2 experience group on the overall connective density (cf. ch. 4.2.3). An ANOVA including the MB group could not be conducted due to a significant inhomogeneity of variance ($F(11, 54)=2,35$, $p<0,019$).[20] Without the MB group the inhomogeneity of variance was barely significant ($F(7, 51)=2,24$, $p=0,046$), however, so that a 2x2x2 ANOVA was conducted with experience group, grade, and sex as influence factors. These results were supplemented with independent samples t-tests to confirm their reliability.

The ANOVA conducted without MB participants showed a (highly) significant main effect not only of experience group ($F(1, 51)=4,62$, $p<0,05$, $\eta_p^2=0,08$) but also of grade ($F(1, 51)=11,61$, $p<0,01$, $\eta_p^2=0,19$). The interaction between grade and experience barely missed significance ($F(1, 51)=3,62$, $p=0,06$, ns). There were no main effect of sex ($F(1, 51)=0,21$, ns) and no other interaction effects.

An independent samples t-test including the MB group confirmed the statistically significant difference between first and fourth grade connective density ($t(45,3)=2,49$, $p<0,05$). The overall variance explained by the grade effect is only 12% ($\eta^2=0,12$), however, which indicates that other factors are (more) important. Further t-tests showed that with regard to L2 experience the statistically significant difference between first and fourth grade connective density is mainly attributable to the mono group ($t(14,8)=3,17$, $p<0,01$), where it explains 40% of the variance in results ($\eta^2=0,40$), even if a trend towards significance was found for the bili group ($t(29,3)=1,74$, $p=0,09$, ns).[21] With regard to participants' sex the difference between first and fourth grade results was statistically significant only for the female group ($t(29,5)=2,65$, $p<0,05$), where it explained 19% of the variance ($\eta^2=0,19$), while the male group's results did not differ significantly ($t(14,7)=0,59$, ns).

T-tests conducted for differences between groups in first and fourth grade found a statistically significant difference between mono and bili first graders' connective density ($t(29)=2,18$, $p<0,05$) with experience group explaining 14% of the variance in results in that grade ($\eta^2=0,14$). No difference between the mono and bili participants was found, on the other hand, in fourth grade ($t(26)=0,16$, ns). With regard to male

[20] Cf. ch. 4.2.3.
[21] The MB group was again excluded since it is too small for any statistical analysis when split up into first (N=3) and fourth graders (N=4).

and female participants t-tests showed no statistically significant differences in either first (t(32)=1,26, ns) or fourth grade (t(30)=0,03, ns).

To sum up, the statistical results show that grade is an important factor for variance in connective density. This importance does not, however, apply to the same degree to all subgroups under investigation. At the same time the statistical analysis showed that L2 preschool experience (mono vs. bili) is a significant factor in explaining variation in connective density results. The influence of the factor experience group is, however, limited to first grade.

6.2.2.5 Summary: Connective density

The present section aimed at answering the question how often participants use connective ties and whether there are any differences attributable to grade, sex or L2 preschool experience. First of all, it was found that almost all participants produced connectives. That is, all fourth graders and 94% of first graders (32 of 34) produced at least some connective ties; only two first graders (6%) did not use any connectives. At the same time participants' interindividual variation in both grades was larger than that obtained for overall cohesive density and reference density (the standard deviation was 0,29 or 60% of the mean in first grade and 0,12 or 19% of the mean in fourth grade).

With respect to an influence of grade it was found that participants' mean connective density differed statistically significantly between first and fourth grade, even if the effect size was comparatively low (grade explained 12% of the variance in results). Thus, first graders produced a mean number of 0,48 connective ties per clause, i.e. they connected roughly every two clauses by a connective, while fourth graders produced an average of 0,62 connectives per clause, i.e. they connected every two clauses with a connective and produced approximately one additional connective in 10 clauses. However, the effect of grade was qualified by sex and experience group. That is, grade significantly affected mono and female results, where it explained 40% and 19% of the variance in results, respectively. The observed increase in bili as well as male results, on the other hand, was not statistically significant. Grade also influenced the interindividual variation in results. Thus, participants' results became far more homogeneous from first to fourth grade (this general development applied to all but the small MB group).

Sex did not have any additional significant effect on connective density, even if all male but not all female first graders produced connective ties. Participants' L2 preschool experience, on the other hand, was found to also exercise a significant observable as well as statistical influence on first grade results. That is, bili first graders produced a significantly higher amount of connective ties per clause (0,53) than mono first graders (0,32), even if the effect size was quite small (experience group explained 14% of the variance in results). At the same time all bili but not all mono first graders produced connective ties and the mono group's results had a much larger interindividual variation (97% vs. 43%). By the end of fourth grade these differences had disappeared, however, so that mono and bili group produced an almost identical number of connective ties (0,62 vs. 0,63) and showed the same degree of variation in results (21%).

The distributions of participants' individual connective density scores showed an overlap between first and fourth grade results for all subgroups. That is, a substantial number of first graders produced a connective density which could just as well have been produced by a fourth grader. The distributions also showed that the main distinction between first and fourth grade was an increase in participants' lower scoring ranges. These findings were again influenced by L2 preschool experience (the difference between grades was more pronounced for mono than for bili participants) but they were not influenced by sex.

To sum up, four general trends were found with respect to participants' connective density, which resemble the ones found for overall cohesive density: A (statistically significant) increase in mean density, a decrease in interindividual variation, an overlap between participants' first and fourth grade results and an increase in the lower scores from first to fourth grade paired with relative stability in the upper ranges. As discussed, these trends were influenced to different degrees by sex and experience group, which interact in several ways so that neither grade nor gender nor experience group alone is able to explain a large amount of the variation in results.

6.2.3 Substitution and ellipsis

As described in ch. 2.4, substitutions and ellipses are two closely related processes and thus they will be covered in the same section. Consequently, the present section aims at answering the question how often participants use ties by substitution or

ellipsis and whether there are any differences attributable to grade, sex or L2 preschool experience.

6.2.3.1 Substitutions

Ties by substitution occurred only marginally in the data. None of the first graders used any such ties and only four of 32 fourth graders (13%) did so. C5-G4-3 (male, bili) produced one substitution, C5-G4-6 (female, bili) two, C5-G4-19 (male, mono) one and C5-G4-20 (female, bili) another one. These five substitutions correspond to a mean number of 0,002 substitutions per clause in fourth grade, i.e. merely two substitutions in 1000 clauses. Due to their extremely limited use substitutions were not investigated further with respect to effects of sex or experience group and they were not submitted to any statistical test. That is, the analysis of substitutions found that the number of ties by substitution increased from first to fourth grade, but that their overall use remained marginal.

6.2.3.2 Overall ellipsis density

Tab. 6.10 gives the descriptive statistics for participants' ellipsis density in first and fourth grade.

	Ellipsis density (No. of ties by ellipsis per clause)					
	Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	34	,09	,08	,07	0	,38
Grade 4	32	,14	,05	,15	,03	,23
Total	66	,11	,07	,11	0	,38

Tab. 6.10 Descriptive statistics ellipsis density by grade

Not all first graders produce ties by ellipsis[22] – this is evident from the minimum score (0) – and altogether the number of ellipses produced is quite low (92 in a total of 1041 clauses). First graders' mean ellipsis density is 0,09, i.e. they produce roughly one ellipsis in 10 clauses. However, the median in first grade (0,07) is even lower than the mean, which indicates that the mean may be slightly misleading as a measure of central tendency – this question will be addressed again together with the distribution of participants' individual ellipsis density scores. At the same time first graders' interindividual variation is extremely high. This is indicated by a standard

[22] The terms 'ellipsis and 'tie by ellipsis' will be used synonymously in the following, since only ellipses functioning cohesively were included in the analysis.

deviation of 0,08, i.e. a variability ratio of 89% in relation to the mean, as well as a range of 0,38 between the minimum (0) and maximum ellipsis density (0,38) obtained in first grade.

Participants' mean ellipsis density increases comparatively strongly (by 150%) from first to fourth grade so that the mean density rises to 0,14 or approximately 3 ellipses in 20 clauses. Even if the mean number of ellipses is still low, e.g. compared to the results obtained for referential ties, the minimum score (0,03) indicates that all fourth graders produce at least some of them. This change also contributes to the decrease in interindividual variation from first to fourth grade. Thus, the standard deviation drops from 0,08 to 0,05 or from 89% to 36% of the mean. At the same time the range from minimum (0,03) to maximum (0,23) drops to 0,2 in fourth as opposed to 0,38 in first grade.

Fig. 6.21 shows the distribution of participants' individual ellipsis density scores in increasing order as a function of grade; the corresponding values are given in Tab. 10.1 and Tab. 10.2 in the Appendix (ch. 10.1).

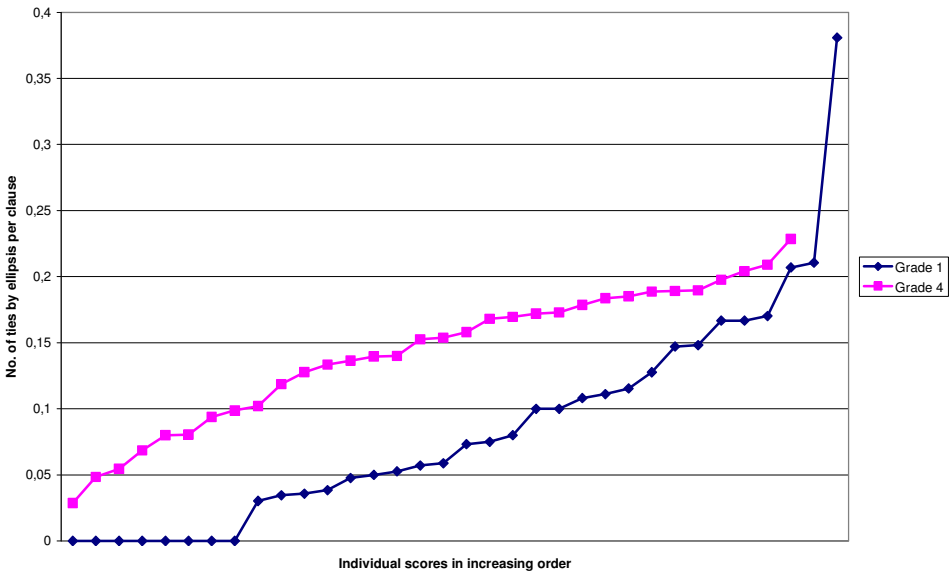

Fig. 6.21 Distribution of participants' individual ellipsis density scores by grade

The first grade curve shows that 24% of first graders (8 of 34) do not produce any ellipses. At the same time it is evident that one first grader (C8-G1-2)[23] distorts the results, since she produces almost twice as many ellipses per clause (0,38) as the second highest score obtained by a first grader (0,21). These extremes were already indicated by the difference between mean and median described above and illustrate

[23] A female first grader without L2 preschool experience.

(as well as are largely responsible for) the strong interindividual variation in first grade. The distribution curve also reflects that – as opposed to first graders – all fourth graders realize at least some ellipses. At the same time a large overlap in ellipsis density results is evident, which ranges from 0,03 (fourth grade minimum and lowest first grade result) to 0,23 (fourth grade maximum). All fourth graders score within this range but also 74% of first graders (25 of 34). 24% of first graders (8 of 34) score lower, namely zero, while one first grader (3%, C8-G1-2) scores clearly above in obtaining the highest score of all participants. Thus, the distribution indicates that also for ellipsis density the main change between first and fourth grade affects the first grade lower scores, in this case the disappearance of participants producing no ellipses at all. The range of results from minimum to maximum remains very high, though. That is, fourth grade results start with a higher minimum, but it is not possible to generally associate a high ellipsis density with fourth rather than first graders.

To sum up, all fourth graders produced ties by ellipsis, while not all first graders did so. At the same time the mean number of ties increased from first to fourth grade, even if their number is comparatively low in both grades. A strong decrease in interindividual variation accompanied the increase in mean ellipsis density. Nevertheless, the variation in results was found to remain high even in fourth grade. The distribution of participants' individual scores showed a large overlap between first and fourth grade and that the main difference between the two grades involved a reduction in the lower scoring ranges, i.e. a higher minimum score in fourth grade caused by the disappearance of zero scores.

6.2.3.3 Ellipsis density by sex

Tab. 6.11 gives the ellipsis density results obtained by male and female participants in first and fourth grade.

		Ellipsis density (No. of ties by ellipsis per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Male	12	,06	,06	,05	0	,17
	Female	22	,10	,09	,08	0	,38
	Total	34	,09	,08	,07	0	,38
Grade 4	Male	10	,15	,06	,18	,05	,21
	Female	22	,14	,05	,15	,03	,23
	Total	32	,14	,05	,15	,03	,23

Tab. 6.11 Descriptive statistics ellipsis density by sex and grade

Neither all male nor all female first graders produce an ellipsis; both groups' minimum score is zero. While both groups scarcely use ties by ellipsis, female first graders have a marginally higher mean ellipsis density with 0,1, as opposed to male first graders, whose mean density is 0,06. That is, female first graders produce one ellipsis in 10 clauses, while male first graders produce approximately one ellipsis in 20 clauses. At the same time both groups' interindividual variation is very high. Thus, male results have a standard deviation of 0,06, which corresponds to a full 100% of the mean, while the female standard deviation is 0,09, which corresponds to a variability ratio of 90%. Similarly, male participants' first grade results have a range of 0,17 from minimum (0) to maximum score (0,17); the female range (0,38) from minimum (0) to maximum (0,38) is even higher.

Fig. 6.22 gives the distribution of male and female participants' individual ellipsis density scores in first grade in increasing order; the corresponding numerical values are given in Tab. 10.3 and Tab. 10.4 in the Appendix (ch. 10.2).

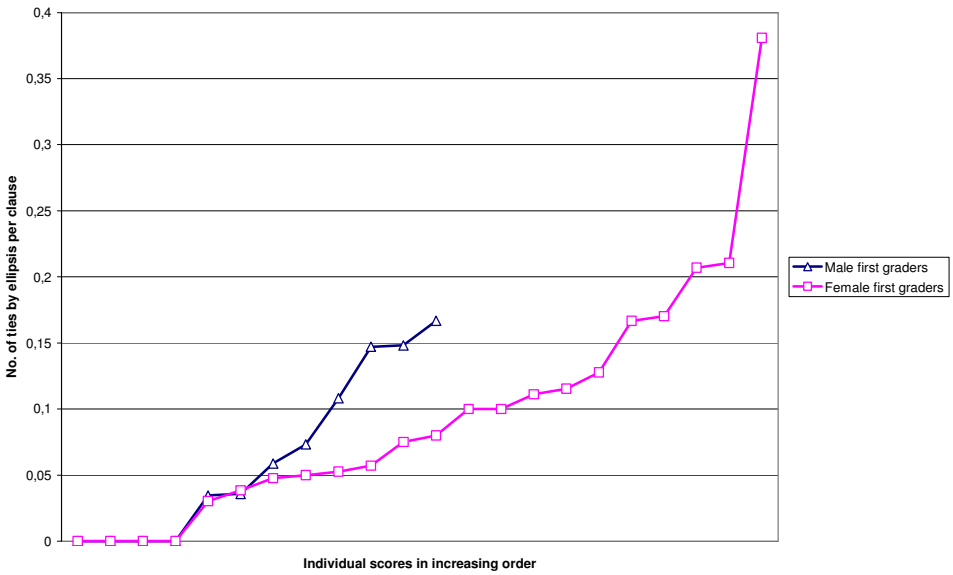

Fig. 6.22 Distribution of participants' individual ellipsis density scores by sex in first grade

The distribution of participants' individual scores shows that 18% of female (4 of 22) and 33% of male first graders (4 of 11) do not produce any ellipses. At the same time the steepness of the curve illustrates the large interindividual variation in both groups of first graders. In spite of the large interindividual variation within the two groups, however, there is a large overlap of scores between male and female first graders, which ranges from 0,03 (both groups' lowest result) to 0,17 (male maximum). 67% of male (8 of 12) and 68% of female first graders (15 of 22) score within this range. Besides the four males (33%) and four females (18%) which do not use any ellipses, 14% of female first graders (3 of 22) score above the overlap; this includes C8-G1-2, who needs to be considered a statistical outlier due to her extreme value.[24]

Both groups' mean ellipsis density increases from first to fourth grade. Male participants' mean density rises a little more (+0,09) than females' (+0,04) so that the two groups' fourth grade results are almost identical: Male (0,15) and female mean ellipsis density (0,14) differ only by one ellipsis in 100 clauses. At the same time both sexes produce at least some ellipses (male minimum 0,05, female minimum 0,03). While their mean ellipsis density increases, both groups' interindividual variation decreases and they become even more similar in this respect. Thus, females' standard deviation drops from 0,09 in first to 0,05 in fourth grade or rather from 90% to 36% of the mean. The male standard deviation remains the same (0,06) but in

[24] If the female mean density is calculated without C8-G1-2, for example, it amounts to 0,08 and is thus clearly closer to the female median – as well as to the male mean.

relation to the mean it decreases from a variation of 100% to 40%. Males' scoring range decreases only marginally (-0,01) to 0,16, while the female scoring range drops by 0,18 from 0,38 in first to 0,2 in fourth grade – a development that is strongly influenced by statistical outlier C8-G1-2, though. Without this extreme value, the female scoring range would also be only marginally different between first (0,21) and fourth grade (0,2).

Fig. 6.23 shows the distribution of participants' individual ellipsis density scores as a function of sex and grade and in increasing order, i.e. on the one hand the differences and/or similarities between male and female participants in first and fourth grade and on the other hand the differences between the two sexes' first and fourth grade results. The corresponding numerical values are given in Tab. 10.3 to Tab. 10.6 in the Appendix (ch. 10.2).

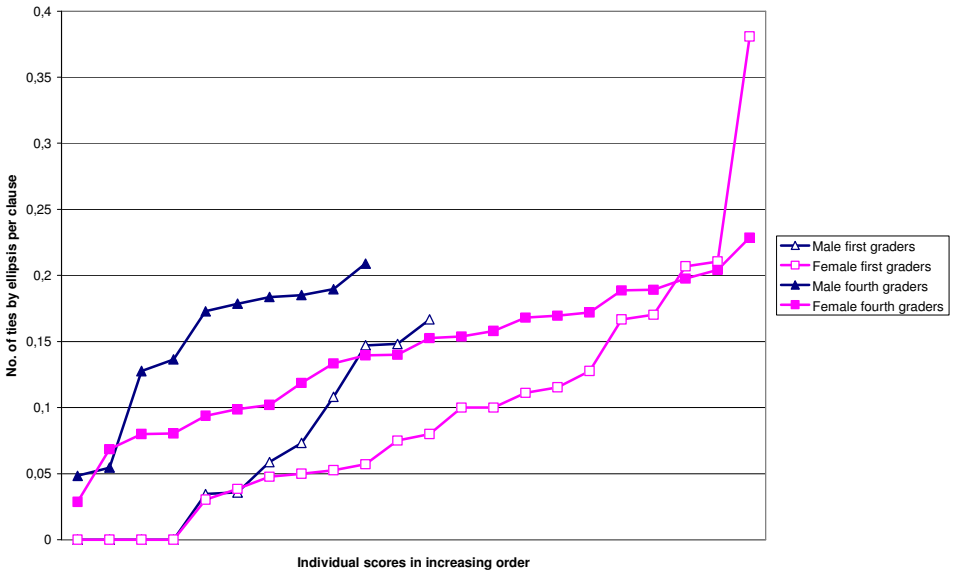

Fig. 6.23 Distribution of participants' individual ellipsis density scores by sex and grade

There is a large overlap of male and female ellipsis density scores in fourth grade, which ranges from 0,05 (male minimum) to 0,21 (male maximum). This range includes all male fourth grade results and 91% of the female fourth graders (20 of 22). One female (5%) obtained a slightly lower score (0,03) and another one scored slightly higher (0,23). That is, there are no significant observable differences in the distribution of ellipsis density scores between male and female fourth graders.

With respect to each sex's first and fourth grade results the following can be observed: Male first and fourth grade results overlap in the range between 0,05 (fourth grade minimum) and 0,17 (first grade maximum). Half of all first graders (50%,

6 of 12) as well as half of all fourth graders (50%, 5 of 10) score within this range. The remaining 50% of male first graders (6 of 12) score below with 33% of them (4) producing no ellipsis at all. The remaining 50% of fourth graders (5 of 10), on the other hand, achieved an ellipsis density above the overlap. Similarly, there is an overlap of female first and fourth grade results in the range between 0,03 (fourth grade minimum and first grade lowest result) and 0,23 (fourth grade maximum). This range includes all female fourth graders' results and 77% of first graders (17 of 22). 18% of female first graders (4 of 22), on the other hand, are not included, since they did not produce any ellipses. One female first grader (5%), namely C8-G1-2, who was already identified as an outlier, scored higher than the overlap.

Thus, male and female distributions do not differ significantly in either grade. Instead, a large similarity can be observed. With respect to each group's first and fourth grade results, female results adhere to the general pattern observed, namely an increase in the lower density scores as opposed to relative stability in the upper ranges. Male results, on the other hand, indicate a more pronounced difference between first and fourth grade ellipsis density and also a fourth grade advantage in the upper ranges. However, this could merely be a group size effect; more male first graders would need to be investigated to confirm this trend.

To sum up, no significant differences between male and female ellipsis density were observed in either grade. Both sexes follow the overall trends identified for ellipsis density, namely an increase in mean density and a decrease in interindividual variation. The distribution of both sexes' results showed a substantial overlap between first and fourth grade scores, even if the difference between the two grades was more pronounced for male participants, and it also showed that the main difference between first and fourth grade is an increase in the lower scores (minimum ellipsis density) just as identified for the overall ellipsis density.

6.2.3.4 Ellipsis density by experience group

Participants' ellipsis density results by experience group and grade are given in Tab. 6.12. As in earlier sections, a main comparison will be made between participants with and without L2 preschool experience (mono vs. bili group).

		Ellipsis density (No. of ties by ellipsis per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Mono	12	,09	,12	,04	0	,38
	Bili	19	,09	,06	,08	0	,21
	MB	3	,06	-	-	,03	,10
	Total	34	,09	,08	,07	0	,38
Grade 4	Mono	12	,14	,05	,16	,05	,20
	Bili	16	,15	,05	,16	,03	,23
	MB	4	,11	-	-	,05	,19
	Total	32	,14	,05	,15	,03	,23

Tab. 6.12 Descriptive statistics ellipsis density by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

As the respective minimum ellipsis densities show, only all three MB first graders produce ellipses; neither all mono nor all bili first graders do so. At the same time mono and bili first graders produce the same mean number of ellipses per clause, namely 0,09 or approximately 1 ellipsis in 10 clauses. The MB mean density is slightly lower with 0,06 or a little more than 1 ellipsis in 20 clauses. As the median shows, however, the distribution of scores seems to be skewed and consequently the mean is somewhat misleading. The median indicates that monos and MBs actually perform lower than the mean suggests and lower than the bili group.

All three experience groups have a high interindividual variation in first grade. The largest variation as measured by the standard deviation is found in the mono group with a standard deviation of 0,12 or 130% of the mean as opposed to a comparatively lower 0,06 or 67% of the mean in the bili group.[25] Similarly, the mono group has the highest range from minimum (0) to maximum (0,38) followed by bilis (from 0 to 0,21).[26] The MB group is the most homogeneous one of the three experience groups as measured by the range of scores (from 0,03 to 0,1).

Fig. 6.24 shows the distribution of mono and bili first graders' individual ellipsis density scores in increasing order; the corresponding values are given in Tab. 10.7 and Tab. 10.8 in the Appendix (ch. 10.3).

[25] The mono results in first grade are somewhat distorted by outlier C8-G1-2, a female mono first grader. If this outlier is eliminated, the mean decreases to 0,06 but the standard deviation (0,07) in relation to the mean is still the highest (117%).
[26] If the outlier is eliminated, the mono range decreases to 0,21.

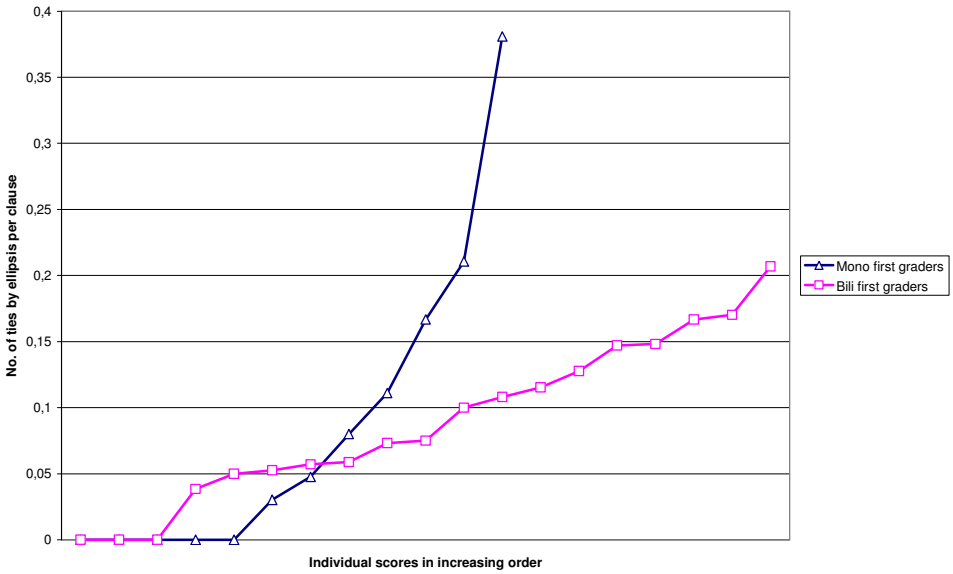

Fig. 6.24 Distribution of participants' individual ellipsis density scores by experience group in first grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience

The curves show that 42% of mono (5 of 12) but only 16% of bili first graders (3 of 19) did not produce ellipses. That is, only 58% of mono (7 of 12) but 84% of bili first graders (16 of 19) used ellipses in their narratives. For the monos and bilis who produced ellipses there is an overlap of results in the range of 0,04 (bili minimum) to 0,21 (bili maximum). All bili first graders who produced ellipses are included in this range, i.e. 84% of the bilis, and 42% of the remaining monos (5 of 12). One mono first grader (8%) scored marginally lower (0,03), while the remaining one (8%, C8-G1-2) was already identified as a statistical outlier, who produces the highest result of all participants in either grade (0,38). That is, the distribution indicates that (a) there are large similarities between the two groups but that (b) bilingual first graders have somewhat of an advantage over mono first graders in terms of the production of ties by ellipsis.

All three experience groups' mean ellipsis density increases almost to the same degree from first to fourth grade (between 0,05 and 0,06).[27] As a result monos and bilis perform again almost identically in their mean density with a difference of merely one ellipsis in 100 clauses (0,14 and 0,15, respectively), while the MB group's mean ellipsis density is slightly lower (0,11). At the same time all fourth graders irrespective of experience group produce at least some ellipses.

[27] Keeping in mind, however, the one extreme value in mono first grade results, the increase of the mono group's mean ellipsis density is actually larger than suggested by the difference in mean between first and fourth grade.

As the mean density increases, the interindividual variation, as measured by the standard deviation, strongly decreases – even if it remains high compared to some of the results described earlier in this chapter. Thus, the mono group's standard deviation in fourth grade amounts to only 0,05 or 36% of the mean, which is almost identical to the bili's standard variation of 0,05 or 33% of the mean. The range of scores, on the other hand, does not drop for all groups: The bili group's scoring range remains fairly stable (0,21 in first and 0,2 in fourth grade), while the mono range decreases very strongly (from 0,38 to 0,15).[28] At the same time the MB group's range doubles (from 0,07 to 0,14). Thus, mono and MB range are almost identical in fourth grade (0,15 and 0,14, respectively), while the bili group's range from minimum to maximum ellipsis density is marginally larger.

Fig. 6.25 shows the distribution of mono and bili participants' individual ellipsis density scores in increasing order as a function of grade, i.e. on the one hand the differences and/ or similarities between mono and bili participants in first and fourth grade and on the other hand the differences between the two experience groups' first and fourth grade results. The corresponding values are given in Tab. 10.7 to Tab. 10.10 in the Appendix (ch. 10.3).

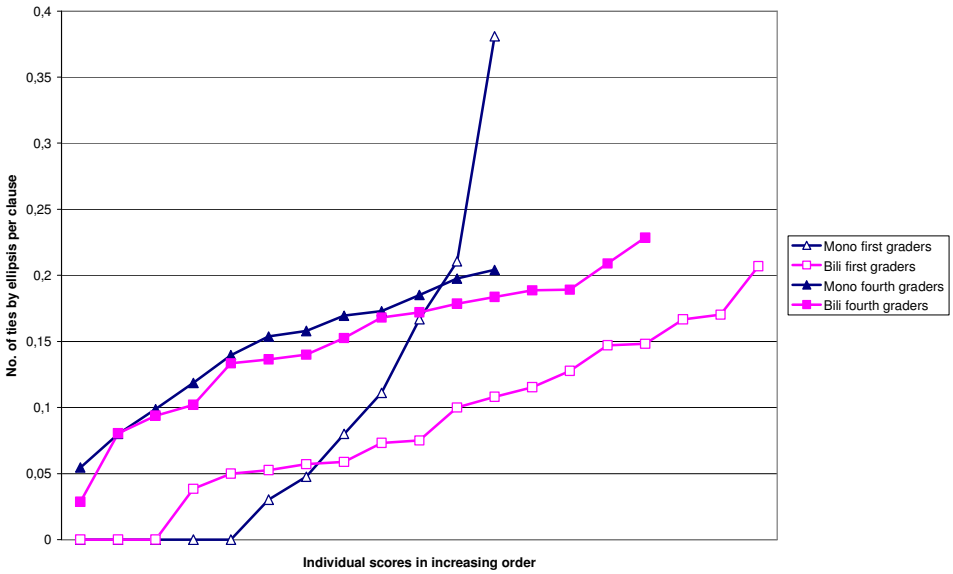

Fig. 6.25 Distribution of participants' individual ellipsis density scores by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

The fourth grade curves show an overlap between mono and bili scores from 0,05 (mono minimum) to 0,2 (mono maximum) ellipses per clause. This range includes all

[28] When disregarding the mono outlier in first grade, however, it decreases only by 0,06 (from 0,21 to 0,15).

mono and 81% of bili fourth graders (13 of 16). Two bili fourth graders (13%) obtained marginally higher ellipsis density scores (0,21 and 0,23), while one bili fourth grader (6%) scored marginally lower than the overlap (0,03). That is, no significant difference is observable between mono and bili fourth graders in terms of the distribution of results.

With respect to mono and bili's first and fourth grade results, the following can be observed: Both groups' curves start out higher in fourth grade (cf. also the respective minimum scores). At the same time there is an overlap of each group's first and fourth grade ellipsis density. Bili results overlap between 0,04 (first grade lowest score) and 0,21 (first grade maximum) with 88% of fourth graders (14 of 16) and 84% of first graders (16 of 19) achieving an ellipsis density within this range. Three first graders (16%) scored lower, namely zero, and one fourth grader (6%) also obtained a marginally lower density (0,03). One last bili fourth grader (6%) scored marginally higher (0,23). Mono first and fourth grade results, on the other hand, overlap between 0,05 (fourth grade minimum and first grade lowest density) and 0,2 (fourth grade maximum). This range includes all fourth graders but only 33% of mono first graders (4 of 12). 50% of first graders (6 of 12) obtained a lower ellipsis density – 42% of which (i.e. 5) scored zero. Two mono first graders (17%) obtained a higher density (0,21 and 0,38). Thus, the comparison between first and fourth grade results shows that both bilis and monos follow the trend of an increase in the lower scores with comparative stability in the higher density results – this phenomenon had already been observed for references and connectives. However, the difference between first and fourth grade results is quite pronounced for the mono group, while the great majority of bili first and fourth graders achieve similar ellipsis density scores.

To sum up, neither all bili nor all mono first graders produced ellipses – all three MB first graders did, however – while in fourth grade all three experience groups produced at least some ellipses. In both grades mono, bili, and MB participants achieved either the same or only marginally different mean ellipsis densities and all three groups followed the general trend of an increase in mean ellipsis densities until the end of fourth grade. However, some differences between experience groups were observed in regard to their interindividual variation. That is, mono first graders' results were much more heterogeneous than those of bili or MB first graders (mainly due to a statistical outlier). Mono and bili group's variation followed the general trend of a decrease in variation from first to fourth grade and by the end of fourth grade

they almost did not differ with respect to interindividual differences (standard deviation and variability ratio). The small group of MB participants, on the other hand, performed somewhat more heterogeneously in fourth than in first grade but (slightly) more homogeneously than monos and bilis (scoring range). The distribution of mono and bili participants' individual ellipsis density scores in first grade showed an advantage of bili over mono first graders, while no differences between groups were found in fourth grade. The distribution also showed a large overlap between each group's first and fourth grade results, which reflected the findings for overall ellipsis density. However, this overlap was less pronounced for mono participants, i.e. their results more clearly differed between first and fourth grade. Thus, distribution and interindividual differences pointed towards an effect of L2 preschool experience on first grade results, while no such effect was observable in the experience groups' mean ellipsis density.

6.2.3.5 Statistical results ellipsis density

General linear model analyses of variance were conducted to statistically test main and interaction effects of grade, sex, and L2 experience group on the overall ellipsis density (cf. ch. 4.2.3). However, both with the MB ($F(11, 54)=2$, $p=0{,}044$) and without the MB group ($F(7, 51)=2{,}29$, $p=0{,}042$) the difference in homogeneity of variance was significant – albeit barely. Consequently, supplementary independent samples t-tests were conducted in order to confirm the ANOVA results' reliability.

An ANOVA including all participants found a statistically significant main effect of grade ($F(1, 54)=5{,}83$, $p<0{,}05$, $\eta_p^2=0{,}1$), which explains 10% of the variance in results, but there were no significant main effects of experience group ($F(2, 54)=1{,}2$, ns) or sex ($F(1, 54)=0{,}38$, ns). Additionally, no significant interaction effects were found.

A second ANOVA conducted without participant C8-G1-2[29], who needs to be considered a statistical outlier (see above), did not yield any different results. Thus, there was again a highly significant main effect of grade ($F(1, 53)=9{,}76$, $p<0{,}01$). The size of the grade effect was larger, however, in this second run ($\eta_p^2=0{,}16$). Sex ($F(1, 53)=0{,}28$, ns) and L2 preschool experience ($F(2, 53)=2{,}15$, ns) were again not significant. Similarly, none of the possible interaction effects was significant, even if

[29] C8-G1-2 is a female first grader without L2 preschool experience.

the interaction between sex and grade showed a trend towards significance ($F(1, 53)=3,1$, $p=0,09$, ns).

Supplementary t-tests including all participants also found a highly significant difference between first and fourth grade ellipsis density ($t(64)=3,35$, $p<0,01$), which accounts for 15% of the variance in ellipsis density results ($\eta^2=0,15$). This significance between first and fourth grade results was also found for male participants ($t(20)=3,24$, $p<0,01$), where it explains 35% of the variance in results ($\eta^2=0,35$), while female participants' results barely missed significance ($t(42)=1,94$, $p=0,059$, ns). With regard to the influence of experience group – the MB group was again excluded due to its small size – a highly significant difference between first and fourth grade was found for the bili group ($t(33)=3,07$, $p<0,01$) with the difference in grade accounting for 22% of variance ($\eta^2=0,22$). The mono group's results, on the other hand, did not differ significantly ($t(14,5)=1,61$, ns) between first and fourth grade.

With regard to differences between the subgroups in first and fourth grade no statistically significant difference was found between mono and bili first ($t(29)=0,11$, ns) or fourth graders ($t(26)=0,24$, ns). Similarly, there was no significant difference between male and female participants in either first ($t(32)=1,09$, ns) or fourth grade ($t(30)=0,46$, ns).

When the same t-tests were conducted without participant C8-G1-2, the difference between first and fourth grade ellipsis density became very highly significant ($t(63)=4,54$, $p<0,001$) and it now accounted for 25% of the variance in ellipsis density results ($\eta^2=0,25$). A t-test without C8-G1-2 also rendered the difference between female first and fourth grade results highly significant ($t(41)=3,16$, $p<0,01$) with grade explaining 20% of the variance ($\eta^2=0,2$). Similarly, the difference between mono first and fourth grade results was now also statistically highly significant ($t(21)=3,31$, $p<0,01$); the factor grade explained 34% of this variance ($\eta^2=0,34$). However, the difference between mono and bili first graders ($t(28)=1,21$, ns) as well as between male and female first graders ($t(31)=0,78$, ns) remained non-significant even though it could also have been distorted by C8-G1-2.

Thus, the statistical analysis shows that – across all subgroups – grade is a highly significant factor for explaining variation in participants' ellipsis density results. Sex or experience group, on the other hand, do not have a significant influence on ellipsis density.

6.2.3.6 Summary: Substitution and ellipsis density

The present section aimed at answering the question how often participants use ties by substitution or ellipsis and whether there are any differences attributable to grade, sex or L2 preschool experience. It was found that only fourth graders produced substitutions and that the overall number of substitutions in the data (5) remained marginal. Due to their marginal occurrence no further analyses of substitutions were conducted. Ellipses were produced more often than substitutions: All fourth graders and 76% of first graders (26 of 34) were found to produce ties by ellipsis. However, the overall number of ellipses produced was also low compared to, for example, participants' reference density, and the interindividual variation in ellipsis density was very high.

Grade was found to have a significant observable and statistical effect on participants' ellipsis density, even if the effect size was relatively low (just as for overall cohesive density and connectives).[30] Thus, first graders achieved a mean ellipsis density of 0,09, i.e. they used roughly one tie by ellipsis in 10 clauses, while by the end of fourth grade the mean density had risen to 0,15, i.e. fourth graders produced approximately 3 ellipses in 20 clauses. While participants' mean density increased, their interindividual variation decreased (from a standard deviation of 0,08 or 89% of the mean in first to 0,05 or 36% of the mean in fourth grade).

The distribution of participants' ellipsis density scores showed a large overlap between first and fourth grade just as found for overall cohesive density, references, and connectives. Thus, 74% of first graders (25 of 34) achieved an ellipsis density that could also have been produced by a fourth grader. A comparison between the first and fourth grade distribution curves showed that the main difference was again an increase in the minimum score from first to fourth grade. However, in the case of ellipsis density this higher minimum was due to a disappearance of participants not producing any ellipses. The minimum density itself remained very low and the variation high. That is, the distribution showed that the statistically significant difference between first and fourth grade ellipsis density is mainly attributable to the first graders not producing ellipses, which explains the relatively low effect size obtained in the statistical analysis.

[30] Grade explained between 10% (ANOVA including all participants) and 25% (independent samples t-test without statistical outlier C8-G1-2) of the variance in ellipsis density.

Sex was not found to have a significant observable or statistical influence on mean ellipsis density or interindividual variation. Instead, male and female participants performed very similarly in first and fourth grade – especially when female outlier C8-G1-2 was disregarded – and they followed the general trends of an increase in mean ellipsis density and a decrease in interindividual variation. The distribution curves showed some distinction between male and female participants in that the difference between first and fourth grade curve was more pronounced for the male group. However, this difference was most likely attributable to a group size effect.

L2 preschool experience was not found to have a significant observable or statistical effect on participants' mean ellipsis density. That is, only marginal differences in mean density were observed among the three experience groups in first and fourth grade. All three groups followed the overall trend of an increase in mean density. However, L2 preschool experience influenced participants' interindividual variation and distribution. Thus, the ellipsis density of participants with L2 preschool experience varied substantially less in first grade than the density of children without such prior L2 experience (the bili standard deviation was 0,06 or 67% in relation to the mean vs. 0,12 or 130% of the mean for the mono group). All three experience groups followed also the general trend of a decrease in interindividual variation from first to fourth grade and by the end of fourth grade the immense difference in variation had disappeared (mono fourth graders had a standard deviation of 0,05 or 36% of the mean, bili fourth graders a standard deviation of 0,05 or 33% of the mean). The distribution of mono and bili individual ellipsis density scores also showed a slight bili advantage in first grade in that 84% of bili (16 of 19) but only 58% of mono first graders (7 of 12) produced an ellipsis. Correspondingly, a comparison of monos' and bilis' first and fourth grade curves showed that the difference between first and fourth grade was more pronounced for mono than for bili participants.

To sum up, it was found that grade significantly influenced ellipsis density. Besides the fact that all fourth but not all first graders produced ellipses, two main developmental trends were identified:

1) a (statistically significant) increase in mean ellipsis density
2) a strong decrease in interindividual variation.

At the same time participants' mean ellipsis density remained comparatively low and the variation comparatively high even in fourth grade. The respective distributions qualified the statistically significant increase in mean ellipsis density, since they showed (a) a significant overlap between first and fourth grade ellipsis density scores and (b) that the difference between first and fourth grade consisted mainly in the elimination of zero ellipsis scores. Sex was not found to have any influence on participants' ellipsis density. L2 preschool experience, on the other hand, was observed to affect first grade results in the sense that first graders with L2 preschool experience had a lower interindividual variation and were more likely to produce ellipses than first graders without L2 preschool experience.

6.2.4 Lexical cohesion

6.2.4.1 Overall lexical density

The present section aims at answering the question how often participants use lexical ties and whether there are any differences attributable to grade, sex or L2 preschool experience. Tab. 6.13 gives the descriptive statistics for participants' lexical density in first and fourth grade.

	Lexical density (No. of lexical ties per clause)					
	Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	34	1,80	,30	1,80	1,22	2,52
Grade 4	32	1,92	,17	1,91	1,58	2,31
Total	66	1,86	,25	1,89	1,22	2,52

Tab. 6.13 Descriptive statistics lexical density by grade

All first graders – as well as all fourth graders – produced lexical ties. At the same time the results obtained for lexical density are the highest of the subcategories of cohesion analyzed and also the ones with the smallest amount of interindividual variation and highest minimum density. Thus, first graders minimally produced 1,22 lexical ties per clause (or approximately five ties in four clauses), and they achieved a mean number of 1,8 lexical ties, i.e. on average they used roughly two lexical cohesive devices per clause. First graders' interindividual differences are comparatively low with a standard deviation of 0,3 or 17% of the mean (cf. the preceding sections), even if the range of their results is quite high with 1,3 lexical ties per clause from minimum (1,22) to maximum score (2,52).

From first to fourth grade participants' mean lexical density increases to 1,92, i.e. fourth graders produce roughly one lexical tie more in 10 clauses than first graders – even if also fourth graders use approximately two lexical ties per clause. While the mean lexical density increases, the interindividual variation decreases. This decrease is evident from an even lower standard deviation, which amounts to 0,17 or 9% of the mean and a reduction in range from 1,3 in first to 0,73 in fourth grade (from a minimum of 1,58 to a maximum of 2,31).

Fig. 6.26 shows the distribution of participants' individual lexical density scores in increasing order as a function of grade; the corresponding values are given in Tab. 10.1 and Tab. 10.2 in the Appendix (ch. 10.1).

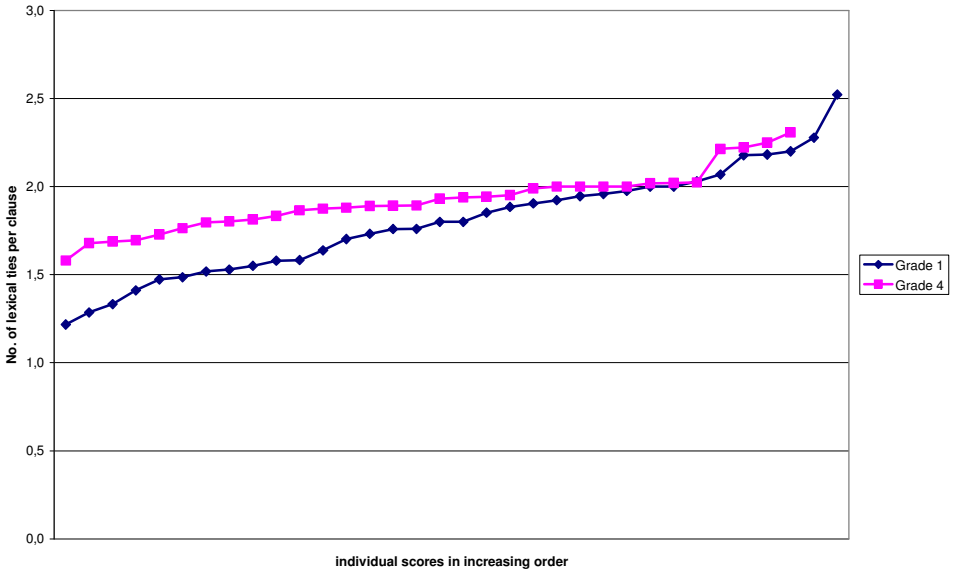

Fig. 6.26 Distribution of participants' individual lexical density scores by grade

There is a large overlap between first and fourth grade lexical density results, which ranges from 1,58 (fourth grade minimum) to 2,31 (fourth grade maximum). All fourth graders achieve a lexical density in this range but also 71% of first graders (24 of 34). 26% of first graders (9 of 34) obtain a lower lexical density.[31] One first grader (3%, C8-G1-16[32]), however, outperforms all other participants by producing 2,52 lexical ties per clause; this first grader needs to be considered a statistical outlier. A comparison of the first and fourth grade curves thus yields a developmental trend which has been described repeatedly for the use of cohesive devices, namely that the main difference between first and fourth grade concerns a reduction of the lower scores, which in this case means an increase in the minimal lexical density.

[31] Up to 0,36 ties per clause lower.
[32] A female first grader without L2 preschool experience.

To sum up, all participants produced lexical ties. At the same time lexical cohesion was found to have the highest density of all subcategories under investigation as well as the highest minimum density. From first to fourth grade participants' mean lexical density increased and their interindividual variation decreased. The distribution of participants' individual density results found that also in the case of lexical density the majority of first graders achieved results that could just as well have been achieved by a fourth grader. Additionally, the distribution showed that the main difference between first and fourth grade was again an increase in participants' minimum score, i.e. an elimination of the lower first grade scores.

6.2.4.2 Lexical density by sex

Tab. 6.14 gives the lexical density results obtained by male and female participants in first and fourth grade.

		Lexical density (No. of lexical ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Male	12	1,89	,25	1,95	1,41	2,28
	Female	22	1,74	,32	1,73	1,22	2,52
	Total	34	1,80	,30	1,80	1,22	2,52
Grade 4	Male	10	1,93	,14	1,92	1,76	2,25
	Female	22	1,92	,19	1,91	1,58	2,31
	Total	32	1,92	,17	1,91	1,58	2,31

Tab. 6.14 Descriptive statistics lexical density by sex and grade

Male first graders produce an average of 1,89 lexical ties per clause, i.e. 0,15 more than female first graders, who use an average of 1,74 lexical ties per clause.[33] At the same time female results are slightly more varied; this is evident from the standard deviation and the range of results. Thus, female first graders' results have a standard deviation of 0,32 or 18% of the mean, while the male group's standard deviation is 0,25 or 13% of the mean. Similarly, female first graders scored on a range of 1,3 from minimum (1,22) to maximum (2,52), as opposed to a male range of 0,87 (from 1,41 to 2,28). The distribution of male and female first graders' individual lexical density scores, which is shown in Fig. 6.27 (the corresponding values are given in

[33] This seems to point to an advantage of male first graders. However, keeping in mind the strong difference in group size, this advantage would have to be confirmed in a larger data set and is thus not considered a robust finding.

Tab. 10.3 and Tab. 10.4 in the Appendix (ch. 10.2), reflects these differences in the range of scores.

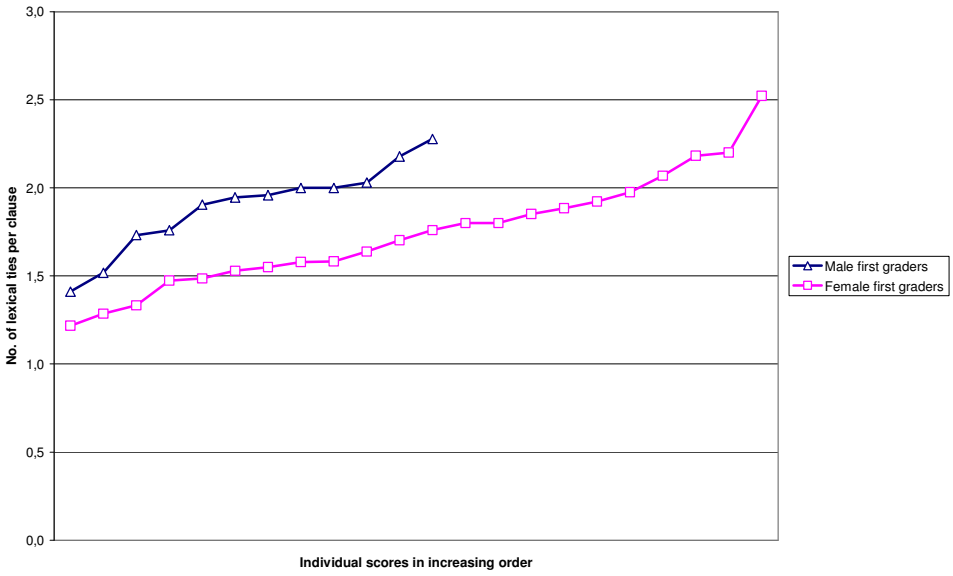

Fig. 6.27 Distribution of participants' individual lexical density scores by sex in first grade

The curve also shows that male and female first graders' results overlap to a large degree. This overlap ranges from 1,41 (male minimum) to 2,28 (male maximum) and includes all male but also the large majority of female first graders (82%, i.e. 18 of 22). Three females (14%) obtained a lower and one (5%) a higher lexical density. This last female (C8-G1-16) simultaneously produced the highest lexical density of all participants (as already described as part of the overall lexical density results). Keeping in mind the different group sizes the distribution thus shows no difference between the two sexes in first grade.

Both male and female lexical density increase from first to fourth grade albeit to different degrees: The mean number of lexical ties increases only marginally for the male group, namely by 0,04 or four ties in 100 clauses. Females' lexical density, on the other hand, increases by 0,18, i.e. by almost two ties in 10 clauses. As a result male and female fourth graders produce virtually the same number of lexical ties, namely 1,93 and 1,92, respectively. That is, both sexes use approximately two lexical ties per clause in fourth grade with a difference of merely one in 100 clauses.

While the mean increases only slightly, both groups' interindividual variation decreases comparatively strongly from first to fourth grade. The male group's standard deviation drops from 0,25, which corresponds to a variability ratio of 13%, to merely 0,14, i.e. a variability ratio of 7%. At the same time males' scoring range

decreases from 0,87 to 0,49 (minimum 1,76, maximum 2,25). The female groups' standard deviation, on the other hand, drops from 0,32 to 0,19, i.e. from 18% to 10% of the mean; their scoring range simultaneously decreases from 1,3 in first to 0,73 in fourth grade (minimum 1,58, maximum 2,31).

Fig. 6.28 shows the distribution of participants' individual lexical density scores in increasing order as a function of sex and grade, i.e. on the one hand the differences and/or similarities between male and female participants in first and fourth grade and on the other hand the differences between the two sexes' first and fourth grade results. The corresponding numerical values are given in Tab. 10.3 to Tab. 10.6 in the Appendix (ch. 10.2).

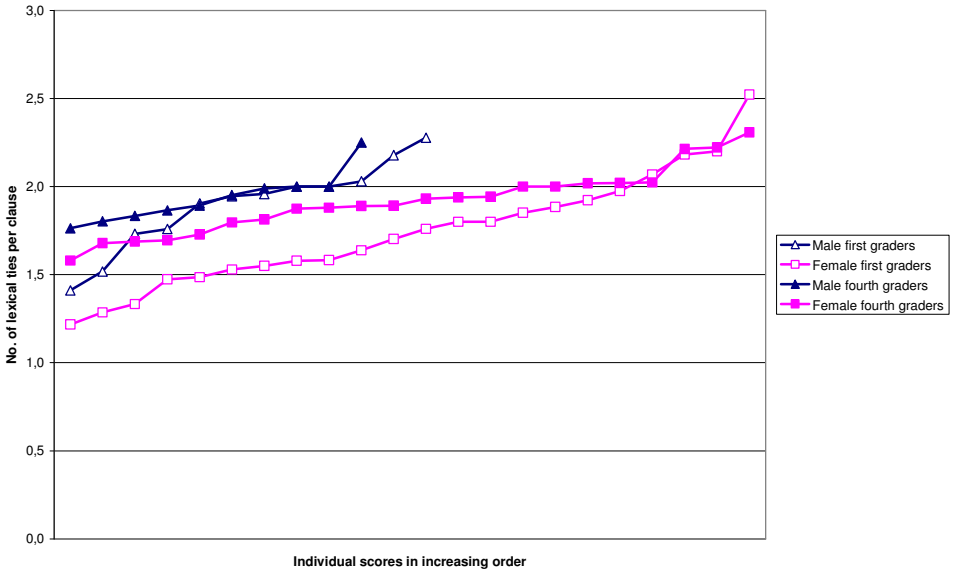

Fig. 6.28 Distribution of participants' individual lexical density scores by sex and grade

A comparison of male and female fourth grade curves shows an overlap between 1,76 (male minimum) and 2,25 lexical ties per clause (male maximum). All male and 73% of female fourth graders (16 of 22) obtained a lexical density within this range. 23% of females (5 of 22), however, scored below and one of them (5%) higher than the overlap. That is, the distribution of individual participants' lexical density scores seems to indicate a slight advantage of male fourth graders. However, this advantage is again more likely attributable to a group size effect and would need to be further investigated with a larger number of participants.

A comparison of either sex's first and fourth grade distribution shows an overlap in scores between the two grades. Male first and fourth grade results overlap in the range of 1,76 (male fourth grade minimum) to 2,25 (male fourth grade maximum). All

male fourth graders achieve a lexical density within this range as compared to 67% of the first graders (8 of 12). 25% of male first graders (3 of 12) scored lower.[34] One male first grader (8%) obtained an even higher lexical density than the upper limit of the overlap (2,28). Female lexical density results, on the other hand, overlap between 1,58 (female fourth grade minimum) and 2,31 (female fourth grade maximum). Again, all fourth graders are included in this range as compared to 64% of first graders (12 of 22). 32% of female first graders (7 of 22) scored lower. Female first grader C8-G1-16 (5%), who was already described as producing the largest number of lexical ties of all participants, scored higher. Thus, a comparison of either sex's first and fourth grade lexical density distribution confirms the general trend of an increase in the lower lexical density scores as opposed to a comparative stability in the upper density scores.

To sum up, no significant differences between male and female participants were found in either grade, even if male first graders achieved a somewhat higher lexical density than female first graders. Both groups followed the general developmental trend identified for lexical density, namely an increase in mean density and a decrease of the interindividual variation to an even lower level than in first grade. Similarly, the distribution of male and female participants' individual lexical density scores followed the findings for the overall distribution, namely a strong overlap between first and fourth grade results and an increase in minimum density scores as the main difference between grades.

6.2.4.3 Lexical density by experience group

Tab. 6.15 gives participants' lexical density results as a function of experience group and grade. As in earlier sections a main comparison will be made between participants with and without L2 preschool experience (mono vs. bili group).

[34] Between 0,03, i.e. only marginally lower, and 0,35, i.e. substantially lower.

		Lexical density (No. of lexical ties per clause)					
		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Mono	12	1,70	,38	1,67	1,22	2,52
	Bili	19	1,82	,25	1,88	1,41	2,28
	MB	3	1,99	-	-	1,80	2,18
	Total	34	1,80	,30	1,80	1,22	2,52
Grade 4	Mono	12	1,87	,18	1,82	1,68	2,31
	Bili	16	1,95	,18	1,94	1,58	2,25
	MB	4	1,96	-	-	1,89	2,00
	Total	32	1,92	,17	1,91	1,58	2,31

Tab. 6.15 Descriptive statistics lexical density by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience and 'MB' children from a monolingual group in a preschool offering also bilingual groups.

Bili first graders produce a mean lexical density of 1,82 ties per clause, i.e. roughly 18 in 10 clauses, while mono first graders use an average of 1,7 lexical ties per clause or 17 in 10 clauses. That is, bili first graders produce on average 0,12 lexical ties per clause or roughly one in 10 clauses more. The small MB group produces the largest number of lexical ties per clause of all three experience groups, namely 1,99. When rounded, however, all three experience groups' mean lexical densities in first grade correspond to two lexical ties per clause.

All three experience groups' results have a comparatively low interindividual variation in first grade: The bili group's results have a standard deviation of 0,25 or 14% of the mean followed by the somewhat more heterogeneous mono group with a standard deviation of 0,38 or 22% of the mean. The mono group also has the largest scoring range (1,3) between minimum (1,22) and maximum lexical density (2,52) followed by the bili group with a range of 0,87 (from 1,41 to 2,28). The MB group's results are the most homogeneous as measured by the scoring range, since they have a range of only 0,28 from minimum (1,8) to maximum lexical density (2,18).

Fig. 6.29 shows the distribution of mono and bili first graders' individual lexical density scores in increasing order. The corresponding numerical values are given in Tab. 10.7 and Tab. 10.8 in the Appendix (ch. 10.3).

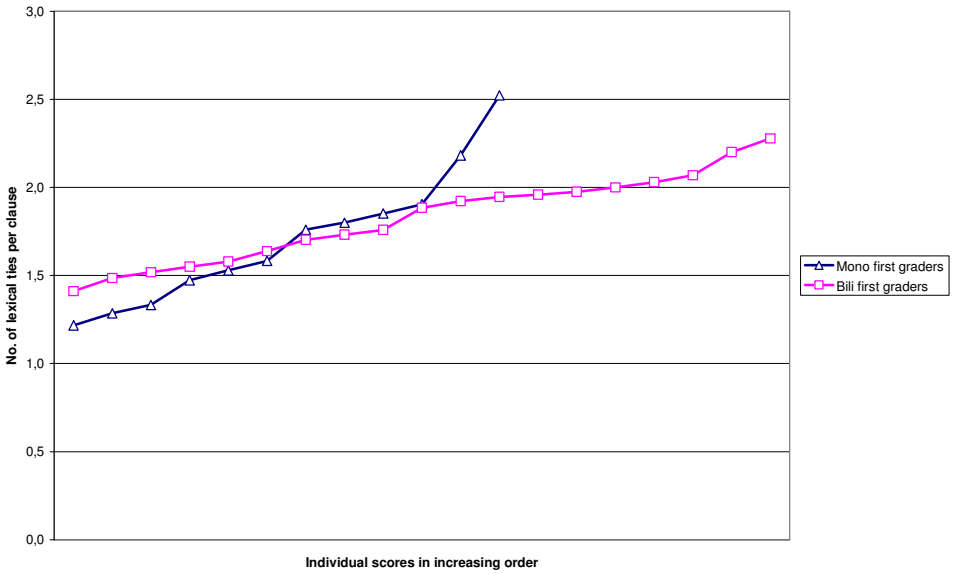

Fig. 6.29 Distribution of participants' individual lexical density scores by experience group in first grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

The curves show an overlap of mono and bili first graders' lexical density between 1,41 (bili minimum) and 2,28 (bili maximum). All 19 bili first graders achieved a lexical density within this range as compared to only 67% of mono first graders (8 of 12). 25% of monos (3 of 12) achieved a lower and one mono first grader (8%, C8-G1-16), who was already described as a statistical outlier, obtained a higher lexical density (2,52). Thus, the first grade distribution seems to indicate a slight advantage of the bili group in that mono first graders are more likely to produce lower lexical density scores than bili first graders.

Mono and bili participants' mean lexical density increases from first to fourth grade (+0,17 and +0,13, respectively), while the MB mean decreases marginally (-0,03). As a result their mean lexical densities become more alike. That is, MB and bili fourth graders produce the largest number of lexical ties per clause with virtually no difference between the two groups (1,96 and 1,95, respectively). Mono fourth graders achieve a marginally lower mean lexical density (1,87) than the other two experience groups with a difference of 0,08 or 8 ties in 100 clauses to the bili result. When rounded, all three experience groups produce again two lexical ties per clause.

All three groups' interindividual variation decreases from first to fourth grade. As a result mono and bili participants become virtually the same in this respect. Thus, bili fourth graders have a standard variation of 0,18 or 9% of the mean and a range of 0,67 between minimum (1,58) and maximum density (2,25). Mono fourth graders also have a standard deviation of 0,18, which corresponds to 10% of their mean, and a

range of 0,63 from minimum (1,68) to maximum (2,31). The small MB group is even less heterogeneous with a scoring range of 0,11 (1,89 to 2).

Fig. 6.30 shows the distribution of mono and bili participants' individual lexical density scores in increasing order, i.e. on the one hand the differences and/ or similarities between mono and bili participants in first and fourth grade and on the other hand the differences between the two experience groups' first and fourth grade results. The corresponding values are given in Tab. 10.7 to Tab. 10.10 in the Appendix (ch. 10.3).

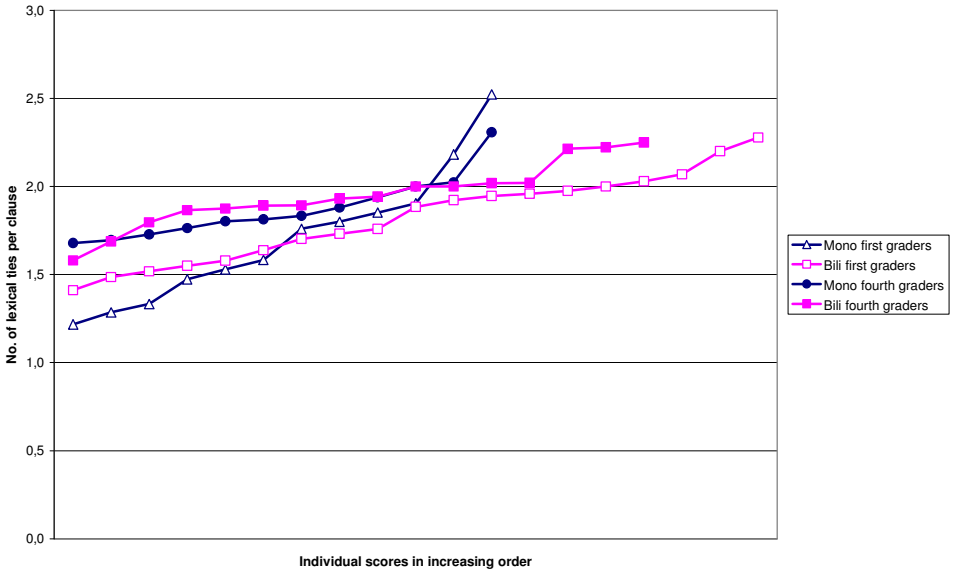

Fig. 6.30 Distribution of participants' individual lexical density scores by experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

There is an overlap between mono and bili fourth grade lexical density scores between 1,68 (mono minimum) and 2,25 (bili maximum). The great majority of both groups' members, namely 92% of monos (11 of 12) and 94% of bilis (15 of 16) achieved a lexical density within this range. Only one mono fourth grader (8%, C5-G4-12) scored higher (2,31) and one bili fourth grader lower (6%, C5-G4-6, 1,58 lexical ties per clause). That is, the distribution of individual participants' lexical density scores shows no difference between mono and bili lexical density scores in fourth grade.

A comparison between either experience group's first and fourth grade results shows similarly large overlaps as they have been described for the overall lexical density as well as for male and female participants. Mono scores overlap between 1,68 (fourth grade minimum) and 2,31 (fourth grade maximum). This range includes all 12 mono fourth graders but only 42% of mono first graders (5 of 12). 50% of mono

first graders (6 of 12) obtain a lexical density below and one mono first grader (8%, C8-G1-16) a density above the overlapping range (the already mentioned statistical outlier). Bili first and fourth grade scores, on the other hand, overlap between 1,58 (fourth grade minimum) and 2,25 (fourth grade maximum). All 16 bili fourth graders obtained a lexical density in this range but also 74% of bili first graders (14 of 19). 21% of bili first graders (4 of 19) achieved a lower[35] and one bili first grader (5%, C1-G1-2) a slightly higher lexical density (2,28). That is, the comparison of each experience group's first and fourth grade curves indicates once more that fourth graders do not necessarily achieve a higher lexical density than first graders. Instead, there is a relatively large overlap as it has been found repeatedly in the analysis of cohesive devices. However, the distribution also indicates that there is a tendency of mono first graders to score less often within the fourth grade range than this is the case for bili first graders.

To sum up, mean lexical density, interindividual variation and distribution taken together seem to indicate a slight advantage of bili over mono first graders. Thus, bili first graders produced a somewhat higher mean density, their results varied less and they were less likely to produce low lexical density scores. The small group of MB first graders was found to outperform both monos and bilis in regard to mean density and variation (scoring range only). Mono and bili participants followed the general trend of an increase in mean lexical density, while the MB group's mean decreased very slightly. All three groups followed the general trend of a decrease in variation, however. As a result of these two developments and distinct increase/ decrease rates any differences attributable to L2 preschool experience were negligible by the end of fourth grade. A comparison of monos' and bilis' first and fourth grade distribution curves showed that both groups followed the general trends of a large overlap between first and fourth grade scores as well as an increase in minimum lexical density as the main difference between first and fourth grade. However, this difference between grades was again found to be more pronounced for the mono group.

6.2.4.4 Statistical results lexical density

General linear model ANOVAs were conducted to statistically test main and interaction effects of grade, sex, and L2 experience group on the overall lexical

[35] Between 0,03 and 0,17 lexical ties per clause lower.

density (cf. ch. 4.2.3). An ANOVA including all participants could not be conducted due to a severe violation of the assumption of homogeneity of variance ($F(11, 54)=2,22$, $p<0,05$). However, when participant C8-G1-16 – a female first grader without L2 preschool experience who can be considered a statistical outlier (see above) – was excluded, the inhomogeneity of variance disappeared ($F(11, 53)=1,95$, ns) and an ANOVA over all other participants could be conducted. This ANOVA did not yield any significant main effects of grade ($F(1, 53)=0,82$, ns), sex ($F(1, 53)=1,94$, ns) or experience group ($F(2, 53)=0,9$, ns) as well as no interaction effects.

A second ANOVA (2x2x2) was conducted without the MB participants, since the MB groups' mean lexical density was found to decrease slightly from first to fourth grade, while all other groups' mean density increased. Without the MB group the difference in homogeneity of variance was barely significant ($F(7, 51)=2,26$, $p=0,044$) so that the ANOVA could be carried out but needed to be supplemented with independent samples t-tests to confirm its reliability. This second ANOVA without the MB group also found no significant main effects of grade ($F(1, 51)=1,07$, ns), sex ($F(1, 51)=0,41$, ns) or experience group ($F(1, 51)=0,62$, ns) as well as no interaction effects.

Supplementary t-tests including all participants showed a significant difference between first and fourth grade lexical density ($t(52,4)=2,1$, $p<0,05$, $\eta^2=0,08$) with grade explaining only 8% of the variance in results, however. With regard to sex this difference between first and fourth grade is mainly attributable to female participants ($t(33,5)=2,18$, $p<0,05$, $\eta^2=0,12$), where grade explains 12% of the variance, while the male group's result failed to reach significance ($t(20)=0,47$, ns). With respect to experience group neither mono ($t(15,5)=1,38$, ns) nor bili results ($t(33)=1,68$, $p=0,103$) differed significantly between first and fourth grade, even if the bili results could still be interpreted as showing a trend towards significance.

T-tests aimed at possible differences between groups in first and fourth grade, respectively, were all non-significant. Thus, no statistically significant difference between mono and bili lexical density results was found for either first ($t(29)=1,1$, ns) or fourth graders ($t(26)=1,12$, ns). At the same time there were also no significant differences between male and female participants in first ($t(32)=1,4$, ns) or fourth grade ($t(30)=0,3$, ns).

Additionally, all t-tests were conducted without C8-G1-16, who could have distorted the results when comparing first and fourth grade as well as first graders by

sex and by experience group. These tests found again a significant difference between first and fourth grade lexical density ($t(53,1)=2,58$, $p<0,05$, $\eta^2=0,11$) and with 11% the variance explainable through the grade effect remained relatively low. Female participants' first and fourth grade results differed again significantly ($t(34,6)=2,9$, $p<0,01$, $\eta^2=0,2$) with grade now explaining a little more of the variance, namely 20%. With respect to experience group there was some change, however, when C8-G1-16 was excluded. Thus, the mono group's first and fourth grade results now differed significantly ($t(16)=2,35$, $p<0,05$, $\eta^2=0,26$) with grade explaining 26% of the variance in results. With regard to the influence of sex and experience group in either grade the statistical results changed only slightly when C8-G1-16 was excluded. The difference between male and female results in first grade now barely missed significance ($t(31)=1,93$, $p=0,063$). Similarly, the difference between mono and bili first graders now showed a trend towards significance ($t(28)=1,91$, $p=0,067$).

To sum up, the statistical analysis showed that grade is a factor influencing lexical density. However, this factor has a comparatively low overall importance and it applies to different subgroups in varying degrees. That is, female and mono results increase significantly from first to fourth grade, while male and bili results do not. Besides this difference sex and L2 preschool experience do not play a statistically significant role in explaining variation in lexical density results.

6.2.4.5 Summary: Lexical density

The present section aimed at answering the question how often participants used lexical ties and whether there were any differences attributable to grade, sex or L2 preschool experience. First of all, it was found that all participants produced lexical ties and that lexical cohesion had the highest density of all subcategories of cohesion under investigation – even in first grade the minimum score was 1,22 lexical ties per clause. That is, all participants produced at least one lexical tie per clause plus an additional one in five clauses. At the same time, participants' interindividual variation was very low in both grades compared to, for example, the variation in ellipsis or connective density.

With respect to an influence of grade participants' lexical density followed two developmental trends, which were already observed for the other subcategories of cohesion, namely an increase in mean density and a decrease in interindividual variation. Participants' mean lexical density increased from 1,8 lexical ties per clause

in first grade to 1,92 ties per clause in fourth grade. That is, first and fourth graders both used approximately two lexical ties per clause, but on average fourth graders produced one lexical tie more in 10 clauses. This difference was also found to be statistically significant, but the effect of grade was quite low (8%),[36] which indicates that other factors are (more) important in explaining variation in mean lexical density. While participants' mean density increased, their interindividual variation decreased from a standard deviation of 0,3 in first to an even lower 0,17 in fourth grade, which corresponds to a decrease in variability ratio from 17% to 9%.

The distribution of participants' individual scores in first and fourth grade showed a similar pattern as found for the other subcategories of cohesion. Thus, a substantial percentage of first graders (71%) achieved a lexical density that could also have been obtained by a fourth grader. At the same time the main difference between first and fourth grade consisted again in an increase in the lower lexical density scores (marked by a relatively strong increase in participants' minimum density).

Sex was not found to have a significant observable or statistical influence on participants' lexical density. Any differences, such as for example a somewhat higher mean density of male first graders (which even showed a trend towards statistical significance when female outlier C8-G1-16 was disregarded), were rather attributable to the large difference in group size. Both sexes followed the general trends of an increase in mean density, a decrease in variation, and a strong overlap between first and fourth grade scores. The statistical analysis qualified these developments to some degree in that only the female group's increase in mean lexical density was statistically significant.

L2 preschool experience was observed to have some influence on participants' lexical density, but this influence was limited to first grade and it was, moreover, not fully confirmed by the statistical analysis, which only showed a trend towards significance (and only once statistical outlier C8-G1-16 had been removed from the analysis). Thus, it was observed that first graders with L2 preschool experience (bilis) achieved a somewhat higher mean lexical density than those without such prior experience (monos; 1,82 vs. 1,7) and that their results varied a little less (the bili group reached a standard deviation of 0,25 or 14% of the mean and the mono group 0,38 or 22% of the mean). Additionally, a comparison between both groups'

[36] 11% if statistical outlier C8-G1-16 was excluded.

distribution curves in first grade showed that mono first graders were a little more likely to obtain a low lexical density than bili first graders (25% of monos, i.e. 3 of 12, achieved a result below a common scoring range). Correspondingly, bili first and fourth grade results overlapped more strongly, while the difference between mono first and fourth grade results was found to be far more pronounced. This observed difference in overlaps was in turn confirmed by the statistical analysis (without C8-G1-16), which showed a statistically significant difference between mono but not between bili first and fourth grade results. However, all three experience groups (mono, bili, MB) followed the general trend of an increase in mean density and a further decrease in variation. Due to distinct increase/decrease rates only marginal differences attributable to L2 preschool experience remained by the end of fourth grade.

To sum up, participants' lexical density followed the same developmental trends found for most other subcategories of cohesion, namely a (statistically significant) increase in mean density and a decrease in variation, even if these trends applied to different degrees to the subgroups under investigation. Sex was not found to influence lexical density results. L2 preschool experience, however, had an observed effect on participants' lexical density in first grade in the sense that children with L2 preschool experience performed slightly better and more homogeneously than children without such prior experience. Lexical density results also showed a strong overlap between first and fourth grade results. This overlap was again influenced by L2 preschool experience in that a few more bili than mono first graders achieved a lexical density in the range of their fourth grade peers.

6.3 The subcategories of cohesion: Contribution to the overall density of participants' stories

6.3.1 Overall contribution

The previous sections described how often participants use each subcategory of cohesive devices, namely references, connectives, substitutions, ellipses and lexical cohesion. A question that remains to be answered, however, is to which degree each subcategory contributes to the overall cohesion of participants' stories, and whether there are any differences in distribution attributable to grade, sex or experience group in regard to this question.

From the results described in the previous sections a clear order of mean densities emerges, which holds true for both first and fourth grade: Lexical ties have the highest mean densities followed by references and connectives. The use of ellipses and substitutions, on the other hand, remains marginal in both grades.

Fig. 6.31 gives the percentage of each subcategory's density relative to the overall cohesive density as a more precise measure for the degree to which references, connectives, substitutions, ellipses and lexical density contribute to overall cohesive density in first and fourth grade.

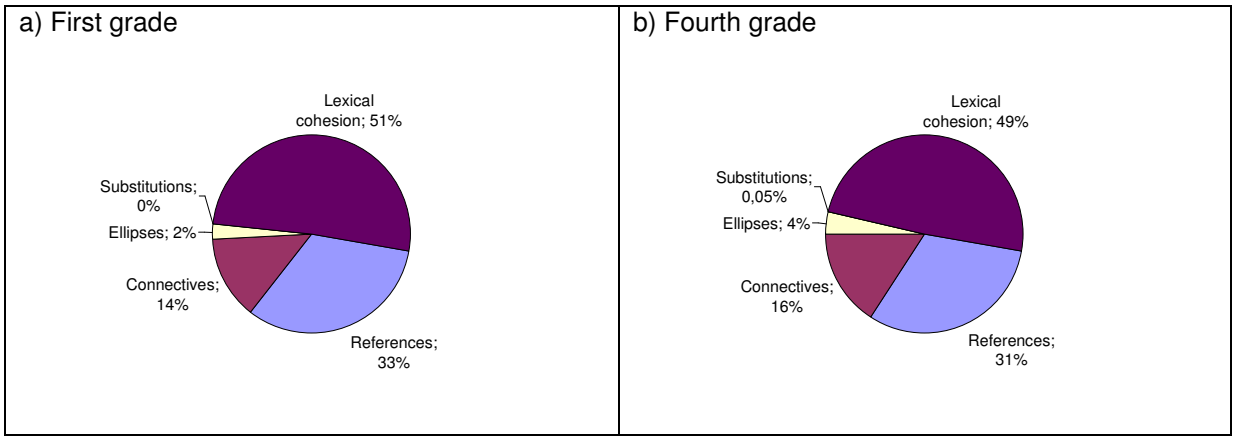

Fig. 6.31 The subcategories of cohesion relative to the overall cohesive density in first and fourth grade

Lexical ties make up more than half (51%) of the overall cohesive density in first grade, while references comprise roughly one third of the overall cohesive density (33%). Connectives constitute 14% of the overall cohesive density in first grade, while ellipses contribute only marginally, i.e. less than 3%, to overall cohesive density (2%). As described earlier, substitutions do not occur at all in first grade.

The contribution of the individual subcategories does not change much from first to fourth grade. Lexical cohesion remains the most important subcategory (49%), followed by references (31%); these two categories' contribution decrease only marginally from grade one to four (-3% and -1%, respectively). The third most important subcategory of cohesion are again connectives (16%), the contribution of which increases very slightly (+2%). As in first grade, ellipses contribute only to a small degree to overall cohesion in fourth grade (4%), even if their contribution has doubled since first grade (+2%). Finally, even though some substitutions were produced in fourth grade, their number is still so small that substitutions account for under 1% of the fourth grade cohesion (0,05%).

6.3.2 Contribution by sex

Fig. 6.32 shows the percentage to which the individual subcategories contribute to male and female overall cohesive density in first and fourth grade.

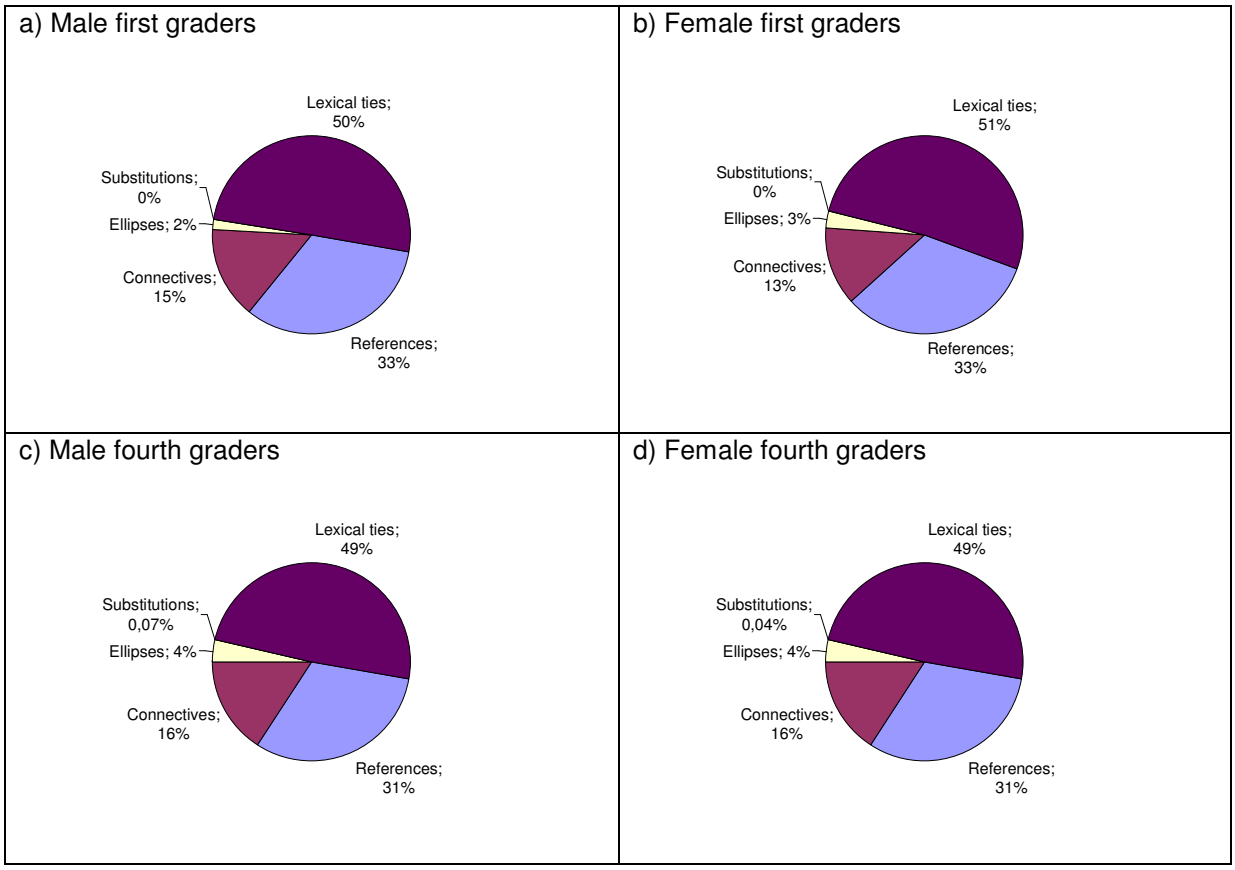

Fig. 6.32 The subcategories of cohesion relative to the overall cohesive density as a function of sex and grade

Both males' and females' distribution of the subcategories is very similar to the overall distribution discussed earlier and the two sexes do not show any significant observable differences in either first or fourth grade: The major bulk of male (50%) as well as female (51%) cohesive density is made up of lexical ties. The next highest contribution is made by references in both male (33%) and female data (33%) followed by connectives in again both male (15%) and female data (13%). Ellipses contribute only marginally to both male (2%) and female (3%) overall cohesive density. Thus, the only difference in first grade with respect to each subcategory's contribution is a marginally larger contribution of connective density in the male as compared to a marginally larger contribution of lexical and ellipsis density in the female data.

Male and female distribution of the subcategories change only slightly from first to fourth grade and in fourth grade there are close to no observable differences

between sexes – both groups' distribution corresponds almost exactly to the overall distribution described above. That is, lexical ties (both groups 49%) and references (both 31%) contribute most to overall cohesive density, even if their contribution is marginally lower than in first grade, while that of connectives has risen slightly to 16% (both groups). The contribution of ellipses has also increased very slightly in male as well as female data and now makes up for 4% of the overall cohesive density in both groups. The contribution of substitutions remains negligible for both sexes (females 0,04% and male 0,07%).

6.3.3 Contribution by experience group

Fig. 6.33 shows the percentage to which the subcategories of cohesion contribute to the overall cohesive density of children with monolingual (monos) and bilingual preschool experience (bilis) in first and fourth grade.[37]

[37] As explained in earlier chapters, the number of MB participants in either grade is too small to allow for any generalization of the results. Consequently, a comparison will be made only between the two main experience groups, namely monos and bilis.

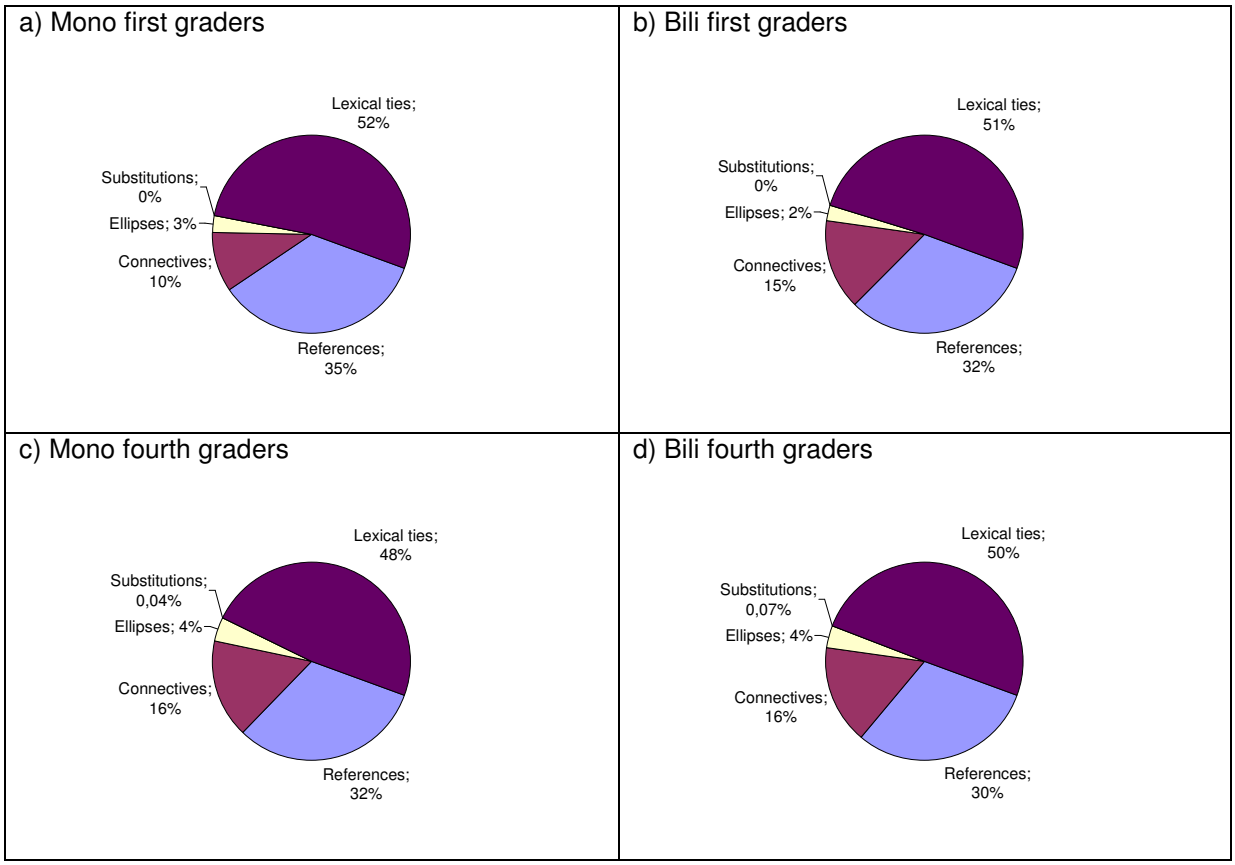

Fig. 6.33 The subcategories of cohesion relative to the overall cohesive density as a function of experience group and grade. Note: 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

The degree to which the different subcategories contribute to overall cohesion in both experience groups' data is similar to the general results described earlier. Thus, in first grade lexical ties are most important in both mono (52%) and bili data (51%), followed by references (35% and 32%, respectively) and connectives (10% and 15%, respectively). Ellipses again make the most marginal contribution to overall cohesion with 3% in the mono and 2% in the bili first graders' data, while substitutions do not occur at all. The first grade results do show some differences, however, between monos and bilis: The distribution of the bili data is influenced by a greater contribution of connectives than in the mono data (a difference of 5%), which in turn is responsible for a marginally lower contribution of lexical ties, references, and ellipses in the bili data. In the lower contribution of connectives monos are also different from the overall first grade distribution described above.

In fourth grade the order of importance of the subcategories remains the same in both groups, even if some changes in the distribution of the subcategories can be observed. That is, lexical ties make up the largest part of the overall cohesive density in fourth grade with 48% in the mono and 50% in the bili data, even if their

contribution has decreased a little in the mono (-4%) and marginally in the bili group (-1%). Similarly, references still make the second largest contribution with 32% in the mono and 30% in the bili data, even if their contribution has also decreased slightly in both mono (-3%) and bili group (-2%). The main difference between mono and bili participants is found in the development of the contribution of connectives. Thus, the contribution of connectives increases substantially in the mono group (+6%) but only slightly in the bili data (+1%) so that by the end of fourth grade the two experience groups do not differ anymore in regard to the contribution of connectives (both 16%). The increase in the contribution of connectives in the mono data also explains its comparatively stronger changes in the contribution of lexical ties and references. In both experience groups the contribution of ellipses remains very low (both 4%), even if it has also increased (monos +1%, bilis +2%), and the contribution of substitutions stays marginal (monos 0,04%, bilis 0,07%). As a result of the changes from first to fourth grade, there are only negligible differences in the distributions of mono and bili fourth graders.

To sum up, a general order of importance was found with respect to the degree to which the subcategories of cohesion contribute to the overall cohesion of participants' stories:

- lexical cohesion (ca. 50%)
- references (ca. 30%)
- connectives (ca. 15%)
- ellipses (< 5%)
- substitutions (< 1%).

This general order is not influenced by grade, sex or L2 preschool experience. The only important differences observed involve the degree of contribution of the individual subcategories in the two experience groups' first grade data. That is, the contribution of connectives is larger in the bili than in the mono first graders' data. By the end of fourth grade, however, this difference has disappeared, i.e. the distribution of the subcategories of cohesion is virtually the same in both mono and bili fourth graders' data and it corresponds closely to the overall distribution in fourth grade.

6.3.4 Qualitative changes

Does that mean there are no overall changes from first to fourth grade besides a slight increase in mean cohesive density? A brief look at the make-up of the three

most important subcategories (lexical cohesion, references, connectives) can help to point towards an answer to this question. Fig. 6.34 shows the distribution of the subcategories of lexical ties in first and fourth grade.

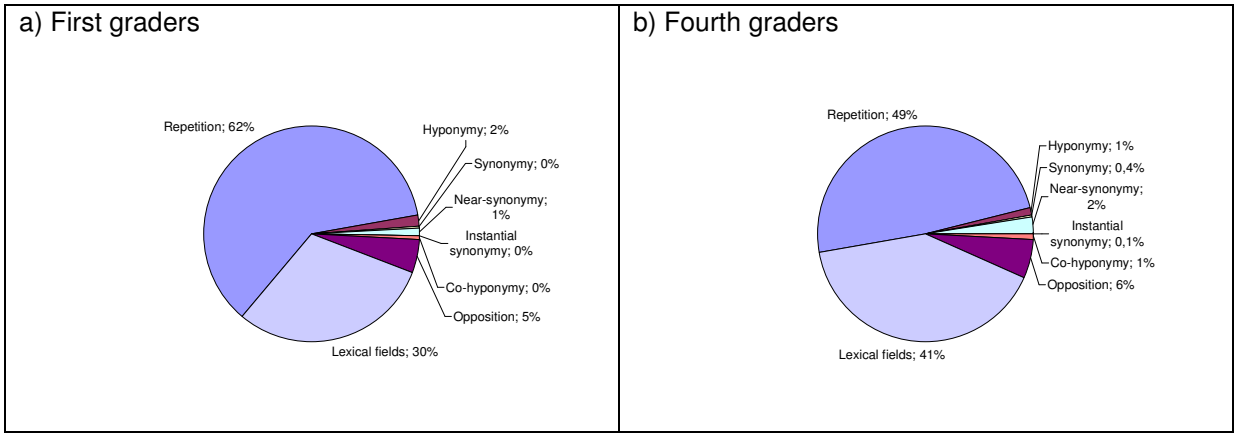

Fig. 6.34 The subcategories of lexical cohesion relative to overall lexical density by grade

The dominant subcategory in either grade are repetitions, followed by lexical fields and then by a fairly marginal use of all other lexical ties. However, a shift in distribution can be observed from first to fourth grade. Thus, the percentage of repetitions decreases by 13%, while the percentage of lexical fields increases by 11% and the percentage of all further subcategories combined also rises slightly (from 8,73 to 10,5%). That is, there is a clear shift in the distribution of lexical cohesion subcategories with the “easiest” subcategory – namely creating ties via simple repetition – decreasing quite substantially, while the subcategories indicative for a stronger lexical diversity increase from first to fourth grade.

A similar but even clearer development from first to fourth grade is evident in the frequencies of the subcategories of reference; this is depicted in Fig. 6.35.

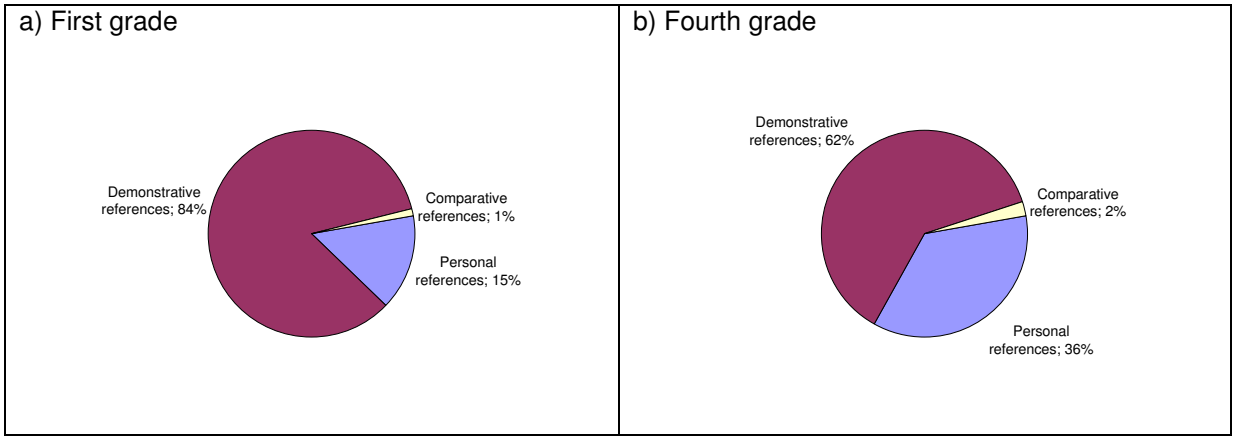

Fig. 6.35 The subcategories of reference relative to overall reference density by grade

Demonstrative references (mainly the definite article) contribute most to the overall reference density in either grade, followed by personal references as the second

most important subcategory. However, the percentage of demonstrative references decreases by 22% from first to fourth grade, while the percentage of personal references increases by 21%. The contribution of comparative references doubles from first to fourth grade but is marginal in both grades. That is, an even clearer qualitative shift in the subcategories of references can be observed with the use of personal references (i.e. pronouns) increasingly replacing demonstrative references (mainly the definite article in constructions such as *the boy*).

A similar qualitative development from first to fourth grade is also evident in the subcategories of connectives; this is shown in Fig. 6.36.

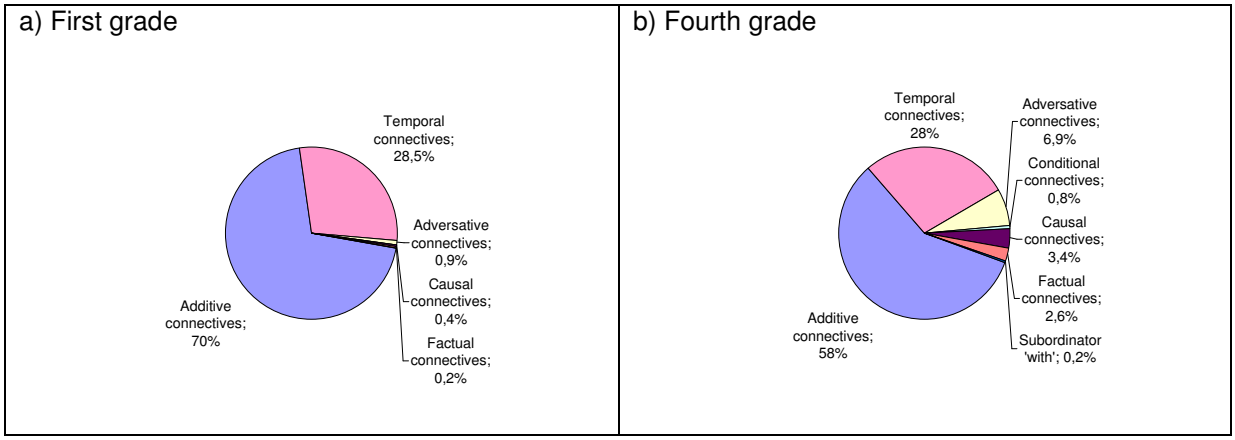

Fig. 6.36 The subcategories of connectives relative to overall connective density by grade

Additive connectives contribute most to connective density both in first and fourth grade, followed by temporal connectives. However, a comparison between first and fourth grade shows that the percentage of additive connectives decreases by 12%. While the percentage of temporal connectives remains fairly stable, it is the connectives expressing more specific relationships that increase. Together, the percentage of the "lesser" connective categories increases from merely 0,15% (adversative, causal and factual) in first to 13,9% in fourth grade. Thus, the percentage of adversative connectives makes up almost 7% in fourth grade (+6%), the one of causal connectives a little more than 3% (+3%) and the one of factual connectives also approximately 3% (+2,4%). Additionally, the use of connectives becomes more diversified, i.e. in fourth grade participants also make (marginal) use of conditional connectives and the subordinator *with*.

6.3.5 Summary: The subcategories' contribution to overall density

To sum up, a general order of degree of contribution to the overall cohesive density of participants' stories was found for the subcategories investigated:

- lexical ties > references > connectives > ellipses > substitutions.

The respective degrees of contribution show almost no changes from first to fourth grade or any influence of sex; only some differences attributable to experience group were found with respect to the contribution of connectives. A more detailed look at the distribution of the subcategories of lexical cohesion, references and connectives, on the other hand, revealed that there are clear qualitative shifts within these subcategories – even if their overall distribution is largely stable from first to fourth grade. Thus, there is a shift from "easier" lexical categories, such as repetition, to categories indicating a more diversified use of vocabulary, a shift from the use of "easier" demonstrative constructions to the use of personal references to refer to story characters, and a shift from "easier" connectives with a general additive or temporal meaning to increasingly encoding relationships between clauses more specifically, e.g. as adversative. These qualitative changes would also need to be investigated with regard to a possible impact of sex or L2 preschool experience. However, this additional analysis lies outside the scope of the present study.

6.4 Cohesion results: Summary

The two previous sections focused on three main questions, namely

- how cohesive participants' stories are as measured by the number of cohesive ties per clause (ch. 6.1)
- whether there are any qualitative differences in regard to the cohesion of participants' stories as measured by the use (ch. 6.2) and degree of contribution (ch. 6.3) of the subcategories of cohesion, i.e. references, connectives, substitutions, ellipses and lexical cohesion
- whether there are any differences attributable to grade, sex and experience group with respect to the first two questions.

Based on the findings of previous studies (cf. ch. 3) the following hypotheses had been put forward with respect to these research questions (cf. ch. .3.6):

(1) Grade/age has a significant influence on the cohesion of participants' stories.

(2) Participants' stories become more cohesive from first to fourth grade as measured by the overall number of cohesive devices.

(3) There are no qualitative differences in cohesion between grades as measured by the frequency order of the subcategories of cohesion (lexical ties > references > connectives > ellipses and substitutions).

(4) Sex does not have a significant influence on the cohesion of participants' stories.

For an easier recapitulation and comparison of results Fig. 6.37 shows the overall mean cohesive density of participants' stories as well as the densities of the individual subcategories by grade as well as their increase from first to fourth grade in one diagram; the corresponding descriptive statistics are given below (Tab. 6.16).

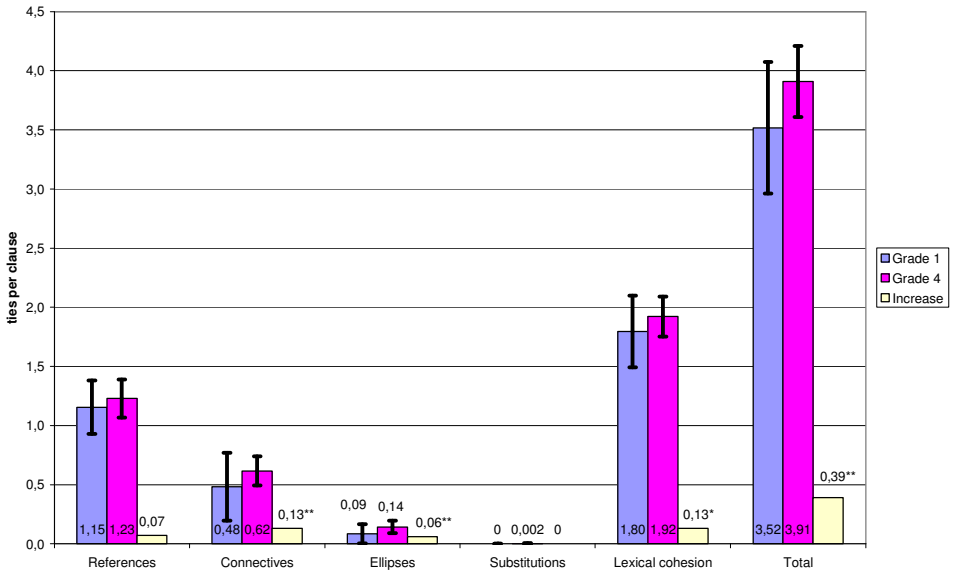

Fig. 6.37 Cohesive densities by grade as well as their increase from first to fourth grade. Note: Statistical significances are not directly comparable, since they are not based on the same tests. The result for connectives is based on an independent samples t-test including all participants; the more conservative ANOVA was only significant when conducted without the MB experience group. The result for ellipses is based on an ANOVA including all participants. Substitutions were not tested, since they only occur in fourth grade. The result for lexical cohesion is based on an independent samples t-test including all participants; the more conservative ANOVA including all participants was not significant. The statistical significance of the overall cohesion of participants' stories is based on an ANOVA excluding the small MB group.

		Total N	Mean	Standard Deviation	Median	Minimum	Maximum
Grade 1	Reference density	34	1,15	,23	1,17	,57	1,65
	Connective density	34	,48	,29	,49	0	,97
	Substitution density	34	0	0	-	0	0
	Ellipsis density	34	,09	,08	,07	0	,38
	Lexical density	34	1,80	,30	1,80	1,22	2,52
	Total cohesive density	34	3,52	,56	3,47	2,17	4,44
Grade 4	Reference density	32	1,23	,16	1,23	,76	1,62
	Connective density	32	,62	,12	,58	,46	,91
	Substitution density	32	,002	,01	-	0	,02
	Ellipsis density	32	,14	,05	,15	,03	,23
	Lexical density	32	1,92	,17	1,91	1,58	2,31
	Total cohesive density	32	3,91	,30	3,89	3,43	4,65

Tab. 6.16 Descriptive statistics cohesive densities by grade

It was found that all participants, i.e. even all first graders, produced at least some cohesive ties; the overall minimum score was approximately two cohesive ties per clause (produced by a first grader). However, not all participants used every subcategory, especially in first grade. Thus, all first graders produced references and lexical ties and most of them also produced connectives. Only some of the first graders, however, used ellipses and none of them produced any substitutions. All fourth graders, on the other hand, produced references, lexical ties, connectives and ellipses. Additionally, at least some of the fourth graders also produced substitutions.

A comparison of the individual subcategories (cf. Fig. 6.37 and ch. 6.3) found a general order of contribution to the overall cohesion of participants' stories, which was not influenced by grade, sex or experience group (in descending order):

- lexical ties > references > connectives > ellipses > (substitutions).

Lexical ties were found to make up about half and references about one third of the overall cohesion, followed by connectives, which contributed about 15%. Ellipses (below 5%) and substitutions (below 1%) were found to be of marginal importance.

At the same time the overall variation in cohesive density results was found to be low compared to the variation found for coherence. However, among the subcategories there were very large differences. Tab. 6.17 summarizes the variability ratios (standard deviations relative to the respective means) by grade, sex and experience group to allow for a better comparison of the results obtained for interindividual variation.

		Total	Male	Female	Bili	Mono
Grade 1	Reference density	20%	16%	21%	13%	28%
	Connective density	60%	49%	66%	43%	97%
	Ellipsis density	89%	100%	90%	67%	130%
	Lexical density	17%	13%	18%	14%	22%
	Total cohesive density	16%	13%	16%	12%	19%
Grade 4	Reference density	13%	11%	14%	14%	13%
	Connective density	19%	18%	21%	21%	21%
	Ellipsis density	36%	40%	36%	33%	36%
	Lexical density	9%	7%	10%	9%	10%
	Total cohesive density	8%	6%	8%	9%	8%

Tab. 6.17 Interindividual variation in cohesive density by grade, sex, and experience group. Note: The interindividual variation is given in terms of the standard deviations as a function of the respective means. Substitutions were not included due to their extremely low number of occurrences. 'Bili' denotes children with bilingual preschool experience, 'mono' those with exclusively monolingual preschool experience.

Thus, a comparison of the individual subcategories' variability ratios as well as their scoring ranges shows the following order of interindividual variation (in decreasing order, cf. Tab. 6.17):

- ellipses > connectives > references > lexical ties.

That is, participants differed most strongly in their use of ellipses, followed by connectives. The results obtained for references and lexical ties, on the other hand, were comparatively homogeneous. The general order of interindividual variation was not found to be influenced by grade, sex or experience group,[38] even if the degree of interindividual variation in first grade was highly influenced by L2 preschool experience in the sense that the results of children with bilingual preschool experience varied far less than those of first graders without prior L2 experience.

Two main developmental trends from first to fourth grade were observed in the data:

- an increase in mean density
- a decrease in interindividual variation.

These developmental trends applied not only to the overall cohesive density but also to all of the subcategories (cf. Fig. 6.37, Tab. 6.16 and Tab. 6.17) and they were not influenced by either sex or experience group. At the same time the rates of increase in mean density varied for the individual subcategories and the following order of

[38] Except for the bili first graders' variation, which was virtually the same for references and lexical ties.

increase rates emerges (fourth grade result minus first grade result, divided by first grade result):

- ellipses (+56%) > connectives (+29%) > lexical ties (+7%) and references (+7%).

That is, the mean densities of references and lexical ties increased only a little from first to fourth grade, while those of connectives, ellipses and substitutions increased very strongly. Even though substitutions were not considered for the order of increase rates, they had the comparatively strongest increase, since they were not produced at all in first grade. However, the statistical analysis qualified the observed order. Thus, overall cohesive density, connectives, ellipses and lexical cohesion differed statistically significantly between first and fourth grade, while this was not the case for references; substitutions could not be tested. At the same time the influence of grade as measured by the respective effect sizes was low compared to the ones found for coherence. Additionally, the overall statistical significance did not always apply to all subgroups and the effect sizes were also influenced somewhat by sex or experience group. Grade explained 40% of the variation in connective density for children without L2 experience, for example, while for the bili group the difference between first and fourth grade connective density only showed a trend towards significance.

With respect to a decrease in interindividual differences (as measured by the respective variability ratio) it was found that the variation of the overall cohesion of participants' stories decreased by 50% from first to fourth grade. At the same time the decrease rates of the subcategories' interindividual variation differed just as strongly as their increase rates in mean density. The following order of decrease rates emerges (first grade variability ratio minus fourth grade ratio, divided by first grade ratio):

- connectives (-68%) > ellipses (-60%) > lexical ties (-47%) > references (-35%).

This general order is not affected by sex or L2 preschool experience. However, participants without L2 preschool experience had much higher decrease rates than the bili participants for the overall cohesive density (-58% vs. -25%) as well as for all the subcategories.

The distribution of participants' individual scores revealed two additional general phenomena:

- an overlap between first and fourth grade scores
- an increase in minimum score from first to fourth grade paired with relative stability in maximum scores.

That is, a large percentage of first graders accomplished density scores which could also have been achieved by a fourth grader. At the same time the major difference between first and fourth grade was found to be a reduction in the lower density scores as opposed to relative stability in the upper density scores. These general observations hold true for all subcategories of cohesion as well as for the two background variables included in the study, i.e. sex and experience group. However, the increase in minimum densities was more pronounced for mono than bili participants.

Sex was not found to have any significant influence on participants' cohesion results; this was true across all subcategories. That is, male and female participants performed very similarly in both grades (no statistically significant differences were found) and they followed the overall developmental trends for mean densities, interindividual variation and distribution. The few differences observed had to be considered unreliable due to a very large difference in group size.

L2 preschool experience was found to somewhat influence cohesion results, but not as significantly as it was the case for coherence (cf. ch. 5). The only stable differences between participants with and without L2 preschool experience were a lower interindividual variation of the bili group in first grade and a stronger overlap of their first and fourth grade scores due to consistently higher minimum scores in first grade. These differences were found for overall cohesion as well as all of the subcategories. Even though bili first graders also had a higher observed mean density for overall cohesion, connectives and lexical cohesion, only the difference in mean connective density was statistically significant. All experience groups largely followed the overall trends of an increase in mean densities, a decrease in interindividual variation and an increase in minimum scores as the main difference between the first and fourth grade distribution of participants' individual results. Due to slightly different increase/ decrease rates and a stronger increase in minimum scores for children without L2 preschool experience any observed or statistically significant differences in mean density and interindividual variation had disappeared by the end of fourth grade.

To sum up, all four hypotheses were confirmed. Grade had an influence on the cohesion of participants' stories in that several developmental trends were found, namely a statistically significant increase in the overall cohesiveness of participants' stories as well as in their use of connectives, ellipses, and lexical ties. The increase in mean density from first to fourth grade was accompanied by a decrease in interindividual variation and an overlap between first and fourth grade scores, which in turn was accompanied by an increase in minimum scores as the main difference between first and fourth grade results. These developments held true for all subcategories of cohesion under investigation, even if increase/decrease rates and overlaps differed for the individual subcategories. No consistent differences attributable to sex were found in the data. Experience group, on the other hand, did have some influence on cohesion results in first grade. That is, the results of first graders with L2 preschool experience had a lower interindividual variation and their first and fourth grade results overlapped more strongly due to higher minimum scores. In regard to participants' mean densities, however, L2 preschool experience only had a statistically significant impact on connective density. By the end of fourth grade any influence of experience group had disappeared.

7 The relationship between (the development of) coherence and cohesion

The present chapter aims at answering the questions whether (1) there is a relationship between coherence and cohesion in participants' narrative discourse and (2) there is a relationship between the development of coherence and cohesion from first to fourth grade. The analysis was limited to investigating linear relationships, however, and any lack of relationship in the results should be interpreted in this light. Thus, several correlation analyses (cf. ch. 4.2.3) were conducted to test whether

- high (low) coherence implies high (low) cohesion and vice versa (positive correlation) or
- high (low) coherence implies low (high) cohesion and vice versa (negative correlation) or
- there is no discoverable linear relationship (no correlation).[1]

The measures submitted to the analyses were participants' total number of narrative components and their narrative index score, which both represent measures of coherence, as well as participants' total number of cohesive devices and overall degree of cohesive density, which both represent measures of cohesion. Consequently, the following correlations are possible in terms of a relationship between (the development of) coherence and cohesion:

- narrative index & cohesive density
- narrative index & total number of cohesive devices
- total number of narrative components & cohesive density
- total number of narrative components & total number of cohesive devices.

For the first question correlation results will be reported separately for first and fourth grade, since it was found that these two groups behave differently and

[1] Analogously the development of coherence and cohesion.

therefore should not be pooled for analysis. Additionally, the results for the mono and the bili group will be reported separately for each grade, since these two groups are also at variance. The MB group could not be submitted to a separate analysis due to its small size, but it was included in the overall analysis in both grades. No distinction by sex was made for the correlation analysis, since sex was not found to be a significant factor in any of the coherence or cohesion results reported in earlier chapters.

While both longitudinal and cross-sectional data were used in addressing the first question, for the second question only the longitudinal data set was included in the analysis. Thus, all of the above measures' increase rates were calculated for participants of the first cohort (fourth grade minus first grade results) and a correlation analysis was conducted among these new increase variables. Only results for the entire first cohort will be reported, since sex was again excluded and experience groups were too small to conduct separate correlation analyses.

Tab. 7.1 shows the correlation matrix for first grade (N=34); Tab. 7.2 and Tab. 7.3 give the correlation matrix for mono (N=12) and bili first graders (N=19), respectively.

	Narrative index	Total number of narrative components	Cohesive density	Total number of cohesive devices
Narrative index	1	,911**	-,092	,196
Total number of narrative components	,911**	1	-,057	,251
Cohesive density	-,092	-,057	1	,576**
Total number of cohesive devices	,196	,251	,576**	1

Tab. 7.1 Correlations in first grade. Note: N=34 for all analyses. ** means that correlation is significant at the 0.01 level (2-tailed).

	Narrative index	Total number of narrative components	Cohesive density	Total number of cohesive devices
Narrative index	1	,980**	-,252	-,001
Total number of narrative components	,980**	1	-,359	-,138
Cohesive density	-,252	-,359	1	,856**
Total number of cohesive devices	-,001	-,138	,856**	1

Tab. 7.2 Correlations in mono first graders' data. Note: N=12 for all analyses. ** means that correlation is significant at the 0.01 level (2-tailed).

	Narrative index	Total number of narrative components	Cohesive density	Total number of cohesive devices
Narrative index	1	,875**	-,211	,278
Total number of narrative components	,875**	1	-,190	,267
Cohesive density	-,211	-,190	1	,237
Total number of cohesive devices	,278	,267	,237	1

Tab. 7.3 Correlations in bili first graders' data. Note: N=19 for all analyses. ** means that correlation is significant at the 0.01 level (2-tailed).

In first grade none of the correlations reflecting a possible relationship between coherence and cohesion is significant when all of the participants are included. The same is true, however, for the analyses conducted separately for mono and bili first graders, i.e. none of the correlations is significant in either of the two experience groups. Thus, the first grade correlation results indicate that there is no linear relationship between any of the coherence and cohesion measures.

Tab. 7.4 shows the correlation matrix for fourth grade (N=32), while Tab. 7.5 and Tab. 7.6 give the correlation matrix for mono (N=12) and bili fourth graders (N=16), respectively.

	Narrative index	Total number of narrative components	Cohesive density	Total number of cohesive devices
Narrative index	1	,404*	-,001	,188
Total number of narrative components	,404*	1	-,076	,612**
Cohesive density	-,001	-,076	1	-,065
Total number of cohesive devices	,188	,612**	-,065	1

Tab. 7.4 Correlations in fourth grade. Note: N=32 for all analyses. * means that correlation is significant at the 0.05 level (2-tailed). ** means that correlation is significant at the 0.01 level (2-tailed).

	Narrative index	Total number of narrative components	Cohesive density	Total number of cohesive devices
Narrative index	-)[2]	-	-	-
Total number of narrative components	-	1	,234	,657*
Cohesive density	-	,234	1	,627*
Total number of cohesive devices	-	,657*	,627*	1

Tab. 7.5 Correlations in mono fourth graders' data. Note: N=12 for all analyses. * means that correlation is significant at the 0.05 level (2-tailed).

	Narrative index	Total number of narrative components	Cohesive density	Total number of cohesive devices
Narrative index	1	,560*	,021	,247
Total number of narrative components	,560*	1	-,234	,606*
Cohesive density	,021	-,234	1	-,474
Total number of cohesive devices	,247	,606*	-,474	1

Tab. 7.6 Correlations in bili fourth graders' data. Note: N=16 for all analyses. * means that correlation is significant at the 0.05 level (2-tailed).

A comparison between the first and fourth grade correlation matrix shows that in fourth grade there is a very highly significant positive correlation between one of the coherence and one of the cohesion measures. Thus, the total number of components correlates very highly significantly with the total number of cohesive devices ($r=0,612$, $p<0,001$, $r^2=0,37$). That is, fourth graders produce either both a high number of components and cohesive devices or they produce a low number of both components and cohesive devices. All other pairs remain without significance, however. The same relationship between coherence and cohesion is found in the mono and bili correlation analyses: The total number of components correlates significantly with the total number of cohesive devices in the mono ($r=0,657$, $p<0,05$, $r^2=0,43$) and equally significantly in the bili data ($r=0,606$, $p<0,05$, $r^2=0,37$). All other pairs did also not yield any significant correlation in either of the two experience groups.

As the correlation matrices showed, total number of components and total number of cohesive devices are related across participants and subgroups in fourth grade. Even more importantly, however, no correlation was found between the

[2] Mono's narrative index results in fourth grade could not be used for the correlation analysis since they do not show any variation and correlation analysis is based on (co-)variance.

narrative index, which focuses specifically on the narrative components indispensable for coherence (cf. ch. 5.3), and any of the two cohesion measures. This phenomenon can be explained by the variation in fourth grade results: As opposed to the interindividual variation in cohesion scores, coherence scores are almost constant; this is reflected by the non-significant correlation.

Tab. 7.7 shows the correlation matrix for the development from first to fourth grade in the longitudinal data set (N=16).

	Increase narrative index	Increase total number of narrative components	Increase cohesive density	Increase total number of cohesive devices
Increase narrative index	1	,676**	-,099	-,297
Increase total number of narrative components	,676**	1	-,049	,321
Increase cohesive density	-,099	-,049	1	,324
Increase total number of cohesive devices	-,297	,321	,324	1

Tab. 7.7 Correlations among increase variables from first to fourth grade. Note: N=16 for all analyses. ** means that correlation is significant at the 0.01 level (2-tailed).

None of the correlations between coherence and cohesion increase measures is significant. That is, both the coherence and the cohesion of participants' stories develop significantly from first to fourth grade (cf. the previous chapters) but their development is not linearly related in individual participants' texts. That is, a strong (weak) increase in coherence does not imply a strong (weak) increase in cohesion and vice versa.

To sum up, no clear-cut answer can be given concerning a relationship between coherence and cohesion. No linear relationship between coherence and cohesion measures was found in first grade – notwithstanding a possible non-linear relation. In fourth grade, however, there is quite a strong linear relationship between the total number of components and the total number of cohesive devices.[3] Changes from first to fourth grade seem to be responsible for this positive correlation. With respect to the development of coherence and cohesion no linear relationship was found.

[3] Additional analyses, which will not be reported in detail, excluded the possibility of a spurious correlation attributable to text length. Text length was found to correlate very highly significantly with the total number of cohesive devices in first ($r=0,946$, $p<0,001$, $r^2=0,89$) as well as fourth grade ($r=0,96$, $p<0,001$, $r^2=0,92$) and it also correlates with the total number of components – significantly in first ($r=0,376$, $p<0,05$, $r^2=0,14$) and very highly significantly in fourth grade ($r=0,593$, $p<0,001$, $r^2=0,35$).

8 General discussion

The present study investigated narrative coherence and cohesion in the L2 story productions of children participating in an immersion program. As stated in the introduction (ch. 1), this study had a threefold aim:

(1) To investigate how linguistic and content organization of (narrative) discourse develop over the four-year duration of an early partial IM program (Goal 1)

(2) To relate this development to participants' cognitive and linguistic development (Goal 2)

(3) To relate the overall results to the effectiveness of the program (Goal 3).

Participants were 66 elementary school children in Germany, who attended an early partial immersion program in which ca. 70% of schooling is conducted in the L2 English (cf. ch. 4). Learner variables taken into account were grade (first vs. fourth, which corresponds to mean age 6;7 and 9;7), sex and L2 preschool experience (especially bilingual vs. monolingual preschool experience). All stories were collected in a picture-elicited oral storytelling task (cf. ch. 4).

In ch. 2 I presented a simplified model of narrative discourse production, which showed two fundamental, interrelated organization principles of texts: Top-down or global and bottom-up or local organization. It was argued that the global organization of (narrative) discourse, i.e. its coherence, is based on an underlying narrative schema. Thus, narrative coherence was defined as a cognitive measure. It was argued further that narrative coherence can be investigated in language with the help of story grammar. Accordingly, narrative coherence was operationalized as narrative components reflecting the underlying story schema in the task material (ch. 4). The number of components, i.e. a quantitative measure, was used to assess the overall coherence of participants' stories. In order to assess qualitative distinctions the individual components' frequency was compared. Additionally, a global narrative index was created, which complements the comparison among individual

components in that it reveals specifically the degree to which a story is organized along the narrative components axiomatic for a global structure.

It was argued further in ch. 2 that local discourse organization is based on the use of linguistic devices to connect clauses into a unified stretch of speech, i.e. cohesion. Consequently, cohesion was defined as a linguistic measure. Cohesion was operationalized via the categories identified by Halliday and Hasan (1976) (cf. ch. 4.2.2). Thus, the density of references, connectives, substitutions, ellipses, and lexical ties was assessed, i.e. their respective mean number per clause, as well as the overall cohesive density of participants' texts, i.e. the total number of cohesive ties per clause. These measures served to assess quantitative differences attributable to grade, sex or experience group. Additionally, it was investigated in how far the individual subcategories contribute to overall cohesive density. This latter measure served to assess qualitative differences.

With respect to the first goal of my study, namely to investigate the coherence and cohesion of participants' narrative productions, I raised several research questions (cf. ch. 2) and put forward corresponding hypotheses (cf. ch. 3). All of these were already reviewed in the respective results sections (ch. 5.4 and 6.4) to which the reader is referred for details. With respect to the second goal two main research questions were posed in ch. 2, which have not yet been taken up again:

1) Is there a cognitive development from grade 1 to 4 as measured by L2 narrative coherence?
2) Is there a linguistic development from grade 1 to 4 as measured by L2 narrative cohesion?

Additionally, the question was raised whether there is a relationship between (the development of) L2 cohesion and coherence and thus, by continuity, between linguistic and cognitive development.

In the following section I will summarize my results with a focus on similarities and differences between (the development of) coherence and cohesion. Then I will review the results obtained in regard to a relationship between these two discourse measures. After that I will discuss two recurring themes of my study: The influence of the three learner variables grade, sex and L2 preschool experience and participants' interindividual variation. Then I will present the answers for questions 1) and 2) (re-) stated above, which derive from the sum of my results, and critically discuss the validity of the constructs of coherence and cohesion. After that some general

limitations of my study will be considered. Finally, the third goal will be taken up again. That is, I will conclude with a brief discussion of the importance of my results with respect to the immersion program at hand as well as immersion education in general.

Coherence and cohesion: Similarities and differences

Since detailed summaries of narrative coherence and cohesion results were given at the end of the respective results chapters (ch. 5.4 and ch. 6.4), the following summary will focus on similarities and differences among results obtained for the two measures.

All participants produced narrative components and cohesive ties. Thus, even first graders were able to produce at least one narrative component and all narrative components were realized by at least one first grader. At the same time first graders also used a minimum of two cohesive ties per clause. However, only references and lexical ties were produced by all first graders. Not all of them, on the other hand, used connectives or ellipses; substitutions were produced exclusively by fourth graders.

Both the narrative coherence and cohesion of participants' stories increased significantly from first to fourth grade, even if the effect of grade was far stronger for coherence than it was for cohesion. At the same time an overlap between first and fourth grade results was observed for both discourse measures. However, this overlap was stronger for overall cohesion (62% of results overlapped). The respective overlaps also showed that for both measures the difference between first and fourth grade scores consisted especially in an increase of the lower first grade scores with relative stability in the upper scoring ranges. This development was, of course, less pronounced for cohesion than for coherence results due to the cohesion results' stronger overlap.

L2 preschool experience affected the overall coherence of participants' stories but not their overall cohesion. At the same time the influence of preschool was limited to first grade. That is, first graders with bilingual preschool experience produced significantly more coherent stories than children without prior experience. Stories' mean cohesive density, on the other hand, did not differ between experience groups. However, L2 preschool experience affected both measures' interindividual variation in that individual results differed less for participants who had attended a bilingual

preschool; this effect was again limited to first grade. At the same time both measures' first and fourth grade scores overlapped more strongly for the experienced group.

Sex was not found to significantly influence overall coherence or cohesion, the respective interindividual differences or the overlap between first and fourth grade scores.

With respect to each measure's subcategories, i.e. individual narrative components and types of cohesive devices, there was also a general tendency for an increase from first to fourth grade: Almost all individual narrative components increased and most often these increases were statistically significant. Connectives, ellipses and lexical ties also increased statistically significantly. However, this was not the case for references, even though the observed results showed an increase. At the same time cohesion results overlapped again between first and fourth grade and this overlap was large for all subcategories (most strongly for references and least for connectives). Thus, the main development concerned the disappearance of the lower first grade scores paired with relative stability in the upper scoring ranges – just as found for overall cohesion.

The frequencies of the individual narrative components differed very strongly in first and less strongly in fourth grade. Similarly, all of the cohesion subcategories followed the trend of a decrease in participants' interindividual variation from first to fourth grade.

An order of frequency was established for narrative components and the cohesion subcategories. For both areas the respective order of frequency was (largely) unaffected by grade. Thus, the frequency order of narrative components indicated an emerging global narrative structure in first grade, which was fully developed by the end of fourth grade (this was confirmed by the narrative index). That is, the major change from first to fourth grade consisted in an increase of each component's frequency, while the general order of frequency did not change significantly. Across grades the frequency order of cohesive devices, on the other hand, indicated that lexical ties and references make the largest contribution to cohesive density (ca. 80%), followed by connectives (ca. 15%). Ellipses and substitutions were found to contribute only marginally. This result confirmed findings of earlier studies (cf. ch. 3).

While sex had again no significant impact on coherence or cohesion results, L2 preschool experience affected both the frequency of individual components and the density of most cohesion categories. However, this preschool effect was again limited to first grade. Thus, all individual components were produced by at least one of the first graders with L2 preschool experience, while this was not the case for the inexperienced group. Additionally, the narrative components' order of frequency indicated that the emerging global organization structure in first grade is limited to the L2 preschool group. With respect to cohesion results the only consistent pattern regarding experience group was again a lower interindividual variation in first grade and a stronger overlap between first and fourth grade scores for children with L2 preschool experience. Only the use of connectives, however, differed significantly between experience groups (experienced first graders had a higher connective density), even if the observed results also showed higher scores for first graders with bilingual preschool experience with respect to lexical cohesion.

The relationship between (the development of) L2 cohesion and coherence

A correlation analysis was conducted among the main measures of coherence and cohesion (total number of narrative components, narrative index score, total number of cohesive devices and overall cohesive density) in order to investigate a relationship between overall coherence and cohesion.

No clear picture emerged. Instead, correlation results differed between grades, even though not between sexes or experience groups. In first grade no statistically significant positive or negative correlation was found between coherence and cohesion. That is, first graders with high (low) coherence scores did not necessarily also produce high (low) cohesion scores (positive correlation). At the same time first graders with high (low) coherence scores did not necessarily produce low (high) cohesion scores (negative correlation).

In fourth grade, on the other hand, coherence and cohesion were related on some measures: There was a positive correlation between the number of components and the number of cohesive devices, i.e. fourth graders producing a high number of components also produced a high number of cohesive devices. This correlation seems to be positively affected by a strongly reduced variation among fourth grade participants, especially with respect to coherence. No correlation was found in fourth grade between cohesion measures and the global narrative index.

However, the index scores' variance was almost non-existent and this may have contributed to a lack of correlation.

What do these results mean for the relationship between coherence and cohesion? The results indicate that global narrative structure and the use of cohesive devices are not per se related, i.e. children producing a fully coherent story as indicated by the index do not necessarily also produce a large number of linguistic devices signalling connections among clauses (absence of a positive correlation). At the same time the results (absence of a negative correlation) also show that more coherent texts are not necessarily associated with fewer cohesive devices, i.e. children do not necessarily compensate fewer cohesive devices with a tighter narrative structure, and at the same time less coherent texts are not necessarily associated with more cohesive devices, i.e. children do not compensate loosely structured narratives with a larger number of cohesive devices. This seems to point towards stylistic reasons for the fourth grade correlation between number of components and cohesive devices, i.e. children preferring more elaborate stories also prefer to produce a higher number of cohesive devices.

While participants' discourse becomes significantly more coherent and cohesive from first to fourth grade, this increase in the two measures is not related, as an analysis of the longitudinal data indicated. That is, the data shows no consistent pattern of development: Children whose increase rates are high (low) for coherence do not necessarily also have high (low) increase rates for cohesion (absence of a positive correlation). Similarly, high (low) increase rates in coherence do not necessarily mean low (high) increase rates in cohesion (absence of a negative correlation). This confirms the conclusions drawn for the relationship between coherence and cohesion.

Several limitations need to be mentioned with respect to these findings. First of all, correlation analysis, which is the standard statistical procedure for testing relationships between variables, only serves to test linear relationships. Further testing would thus be desirable to investigate also the possibility of a non-linear relationship. Secondly, a more detailed look at correlations involving submeasures of coherence (individual narrative components) and especially cohesion (e.g. connectives) would need to be included in future studies (cf. also Bae 2001:74). Finally, it would be desirable to also include coherence and cohesion measures derived from more qualitative analyses, e.g. reference strategies.

Learner variables and their impact: Grade, sex and preschool experience

On the basis of previous research two hypotheses were put forward with respect to the learner variables investigated (cf. ch. 3):

(1) Grade has a significant influence on the narrative coherence and cohesion of participants' stories

(2) Sex has no significant influence on the narrative coherence and cohesion of participants' stories.

No hypothesis was formulated regarding L2 preschool experience.

The results obtained for coherence and cohesion showed a common pattern, which confirms the first hypothesis: Grade is the single most influential factor in explaining variation (cf. the sections above). However, grade explains variation in coherence far better than in cohesion. Thus, the effect sizes were very high for coherence measures, namely ca. 50% for total number of components and global narrative index score and 32% to 77% for the individual narrative components, while the effect sizes for cohesion were comparatively low, namely 12% for overall cohesion and typically between 8% and 12% for the subcategories (without outliers maximally 25%). Furthermore, grade had no statistically significant effect on one of the subcategories of cohesion (references).

So far grade has been treated as one complex influence factor, which incorporates exposure as well as maturational effects. The question arises, however, whether this is feasible. That is, the question needs to be addressed whether maturational effect and exposure effect can be isolated. This question was already touched on in passing in the overview of previous research (cf. ch. 3) and I would like to argue that both effects overlap in my study.

This is especially evident for narrative coherence: As explained in ch. 3, the age range of the present study's participants corresponds to the one in monolingual children where a strong development in coherence is to be expected. There is no readily conceivable reason why this should not also apply to the present study's participants – even if they tell stories in an L2 – since narrative coherence is a cognitive measure. That is, a maturational effect is to be expected. The impact of L2 preschool experience on first grade coherence results, on the other hand, indicates an overlap of both factors – at least if one does not assume that children with L2 preschool experience are per se cognitively more advanced. This latter interpretation

does not hold against the fact, however, that children without bilingual preschool experience are able to catch up until the end of fourth grade. Thus, the substantially weaker performance of first graders without L2 preschool experience is more likely attributable to linguistic deficits. Since neither experience group consistently produces coherent stories in first grade but both do so in fourth grade, this in turn means that development is attributable to a combination of maturational and exposure effect. Consequently, these two factors are indeed inseparable in my study.

No significant influence of sex was found in the analyses. That is, there were no statistically significant differences between male and female participants with respect to cohesion or coherence; this confirms the second hypothesis. Only the observed results indicated some minor differences, e.g. with respect to connectives (ch. 6.2.2.2) and lexical cohesion (ch. 6.2.4.2), but no consistent pattern emerged. The lacking influence of sex is in line with the scarce number of similar studies, which found either no differences or only qualitative differences within subcategories, which were not investigated in the present analysis (cf. ch. 3). It would thus be desirable to conduct a more fine-grained analysis of the subcategories in the future to see whether this yields any more substantial differences attributable to participants' sex.

The impact of L2 preschool experience is limited to first grade, even though the development from first to fourth grade could also be affected by the difference in first grade scores. By the end of fourth grade any differences attributable to experience group have evened out. At the same time the only consistent influence of experience group on both coherence and cohesion scores concerned interindividual variation, which was lower for first graders who had attended a bilingual preschool. Any other effect of experience group was limited mostly to coherence measures, where first graders with L2 preschool experience generally outperform those without prior experience. The effect sizes for this phenomenon were comparatively high. Thus, preschool experience explains between 23% (index score) and 33% (total number of components) of variation in first grade.

The influence of preschool experience on cohesion is far less evident apart from the lower variation for first graders with bilingual preschool experience, which was described above. With respect to overall cohesion, as well as the subcategories, the difference between mono and bili first graders reaches significance only for connectives and here preschool experience explains merely 14% of the variation in results. This is somewhat surprising, since the model outlined in ch. 2 purports

cohesion as a linguistic measure and thus clear differences among groups would be expected for cohesion and much less so for coherence measures. This issue will be taken up again in one of the following sections.

Even grade as the single most important factor in explaining variation leaves about 50% of the variation in coherence unexplained and 88% or more for cohesion results. Thus, other factors seem to influence coherence and especially cohesion in L2 narrative discourse, even if some degree of random variation among participants is allowed for. One of the limitations of the present study is that only the influence of grade, sex, and L2 preschool experience was considered. That is, many potential influence factors were not included, such as socioeconomic status, L1 proficiency, general cognitive abilities, motivation or activities and skills strongly related to storytelling, such as story comprehension or parents' storytelling (cf. also overviews of learner variables, e.g. Larsen-Freeman 1991, Ellis 1994). Any future study on coherence and cohesion would thus need to include as many additional impact factors as possible.

Interindividual variation

Several general phenomena were discovered with respect to interindividual variation as measured by the ratio of standard deviation to mean. First of all, it was found that interindividual variation is influenced by grade. That is, participants' interindividual variation is higher in first than in fourth grade; this pattern agrees with other studies conducted in the Kiel IM project (cf. Wode 2009: 37). An influence of experience group was also found in that (a) interindividual variation is lower for first graders with bilingual preschool experience than those without such prior experience, (b) the interindividual variation of participants without L2 preschool experience decreases more strongly from first to fourth grade and (c) by the end of fourth grade any differences in interindividual variation among experience groups have disappeared. Sex had no influence on interindividual variation, just as it did not significantly affect any of the other results obtained for coherence and cohesion.

Besides the learner variables grade and experience group, however, interindividual variation was also found to differ among the coherence and cohesion measures investigated. First of all, first graders' interindividual variation was far larger for coherence than cohesion: Variation in overall coherence amounted to 65% (ratio of standard deviation to mean) in first grade for the number of components and 46%

for the narrative index, while it was only 16% for overall cohesive density. This order was stable also across experience groups. Interindividual variation decreased far more strongly for coherence measures, though, so that by the end of fourth grade it was quite similar for both coherence and cohesion. Thus, participants' variation in overall coherence amounted to 16% in fourth grade for the number of components, 5% for the narrative index and 8% for overall cohesive density.

As the comparison between number of components and narrative index indicates, differences in interindividual variation between these two coherence measures are mainly attributable to the variety of components produced and much less to differences in global organization. This points to stylistic reasons for the difference in variation between number of components and index score, especially in fourth grade.

The question remains which factors are responsible for the much larger variance in coherence than in cohesion results, especially in first and partly in fourth grade. Possibly, this is an artefact of the method of analysis. That is, participants could have realized none of the narrative components, even if they produced a stretch of speech in relation to the task material. It would be far more difficult, however, to produce a stretch of speech not containing any cohesive ties. It is difficult to imagine, for example, how participants could produce utterances in relation to the task material which do not contain any content words. Any stretch of clauses thus very likely includes, for example, repetitions of a lexeme denoting the main character(s) *boy* or *dog* and, consequently, a certain number of cohesive devices. That is, participants may not be able to avoid producing a minimum number of ties, since certain characters and phenomena are presented repeatedly in the task material.

At the same time interindividual variation in cohesion was strongly dependent on the type of cohesive device. Variability in results was highest for ellipses (89% in first and 36% in fourth grade), followed by connectives (60% vs. 19%), references (20% vs. 13%) and lexical ties (17% vs. 9%). That is, the amount of variation was dependent on grade but not its general order – just as this order was not influenced by experience group, even if first grade results varied less for children with L2 preschool experience.

This could indicate that some cohesive ties are more difficult to produce than others. The use of ellipses, for example, is strongly dependent on the use of complex

syntax. If participants produce only simple sentences – as they are quite likely to do in first grade – the structural prerequisites for ellipses will not be met and thus no ellipses produced. At the same time the use of some types of cohesive devices may be more flexible than that of others. That is, some cohesive ties are artefacts of task and task material (especially lexical ties, cf. above), while others enjoy a much greater stylistic flexibility. This seems to be the case for connectives, for example, the use of which is mostly optional.

Coherence and cohesion: Cognitive and linguistic development

As I outlined in ch. 2 and summarized again at the beginning of the present chapter, narrative coherence, which is based on the application of a narrative schema as a global production plan, was defined as a cognitive measure. The use of cohesive devices, on the other hand, was defined as a linguistic measure based on the use of cohesive devices, which serve to establish local relations among stretches of clauses and to thereby connect theses stretches of clauses into a unified whole. By continuity, the development of narrative coherence and cohesion from first to fourth grade was assumed to reflect participants' cognitive, respectively linguistic development. Thus, the following research questions had been put forward in ch. 2 and were repeated at the beginning of the present chapter:

1) Is there a cognitive development from grade 1 to 4 as measured by L2 narrative coherence?
2) Is there a linguistic development from grade 1 to 4 as measured by L2 narrative cohesion?

As my results showed, both narrative coherence and cohesion increase significantly from first to fourth grade. This in turn means that participants in the immersion program develop cognitively as well as linguistically over the three years between test time 1 and test time 2. However, several questions arise as to the validity of coherence and cohesion as a cognitive, respectively linguistic measure.

Coherence, first of all, does not only show differences between grades but also between experience groups in first grade. That is, participants without L2 preschool experience were found to produce significantly less coherent stories in first but not in fourth grade. Does this mean that first graders with bilingual preschool experience are cognitively more advanced than those without prior experience? Such an interpretation is possible but problematic and it would have to be cross-checked very

carefully with evidence for a narrative schema in the L1 (as well as with further cognitive measures). Rather, the significant difference between experience groups in first grade points to the influence of a linguistic threshold. That is, first graders with bilingual preschool experience have the means to express the narrative schema in the L2 more readily available than those without such prior experience (cf. ch. 3 for an analogous phenomenon in L1 acquisition).

Even though it could similarly be claimed that the large differences in coherence between first and fourth graders are attributable to linguistic difficulties, this does not seem likely for two reasons. First of all, even for monolingual children the major development in global narrative structure has been found to take place within the age range under investigation (cf. the developmental overview in ch. 3). Secondly, neither group consistently structured their narratives according to a global narrative schema in first grade: Significantly more but not all first graders with L2 preschool experience produced a(n) (almost) fully coherent story. At the same time some of the children without prior experience also produced a(n) (almost) fully coherent story.

This in turn means that the usefulness of coherence as a cognitive measure can be obscured by a linguistic threshold. This applies, of course, as much to L2 as to L1 studies, even if the discrepancy between cognitive and linguistic expression may be more salient for L2 speakers. Consequently, coherence can be regarded as a useful measure only for cognitive development over time.

How about cohesion? The degree of cohesive density, i.e. the number of ties per clause, differs significantly between grades. However, the sum of cohesion results casts doubts whether the use of cohesive devices is a straightforward indicator for linguistic development, since (a) not all subcategories develop significantly, (b) interindividual variation is comparatively low but differs strongly among subcategories (cf. the previous section), (c) a general order of frequency applies to the subcategories and their contribution to overall cohesion and (d) no significant differences between experience groups were found except for the use of connectives, even though linguistic differences would be expected at least in first grade. What does this mean for the construct validity of cohesion?

Differences in interindividual variation among cohesion subcategories were already discussed in the previous section. Linguistic complexity and stylistic flexibility were identified as possible reasons for such differences. It was also discussed that

the use of at least some cohesive devices can hardly be avoided if any utterances are produced in relation to task and task material.

The general order of frequency and contribution (cf. also the findings of other studies described in ch. 3) points additionally to structural influences on the use of cohesive devices. That is, clauses minimally have an SV-structure, i.e. they contain a subject, which will minimally be a noun, noun phrase or pronoun, and a verb. Thus, any clause will normally contain at least one word that is a potential lexical tie (the verb). At the same time participants, especially first graders, often fill the subject slot with a full noun phrase. This means that with great likelihood (a) the definite article, i.e. a demonstrative reference, will be used and (b) a noun, i.e. a word that has again the potential of a lexical tie. A typical SV clause will therefore contain at least two potential lexical ties and one potential (demonstrative) reference; this is reflected in my results (cf. ch. 6.2). The use of connectives is, on the other hand, far more flexible, since semantic relationships between clauses are often inherent in their content and do not have to be made explicit. Ellipses, finally, are bound to the use of complex syntax, i.e. they have a strong need for certain structural prerequisites to be fulfilled, and their use is therefore far more limited. Thus, differences in interindividual variation as well as general order of frequency and contribution to overall cohesion point to a limited comparability among subcategories' density.

However, cohesion results raise an even more crucial question: What is a 'cohesive' text and how cohesive does a text need to be? The importance of this question is reflected in my results for the relationship between coherence and cohesion, which showed that both measures are not consistently related and that their development from first to fourth grade is parallel but not related. That is, the degree of cohesion was not found to necessarily correlate positively or negatively with the degree of coherence. The changes found within subcategories (cf. ch. 6.3.4), also tie in with this question, since they showed that the majority of changes from first to fourth grade seems to take place within subcategories and not in overall cohesive density or the density of each subcategory.

In the light of the issues raised above it is possible that Halliday and Hasan's (1976) approach to cohesion obscures more important phenomena, such as more general discourse phenomena being governed by underlying cognitive strategies, e.g. referring to story characters. The linguistic implementation of reference strategies, for example, would cut across Halliday and Hasan's categories of

personal reference, ellipses (zero anaphora), demonstrative references and lexical ties. Relatedly, the use of connectives may be governed by a "difficulty of inference" strategy, which calls for the use of a connective only if the storyteller believes that the relationship between two clauses is not easily inferable.

To sum up, the construct validity of cohesion within the framework of Halliday and Hasan (1976) is not entirely robust. What does this mean for discourse as the result of top-down and bottom-up organization as outlined in the simplified model of narrative discourse production presented in chapter 2.1? The two organizational structures identified for texts, top-down and bottom-up, were only partly confirmed in the way they had been defined. Rather, it may be necessary to redefine top-down organization as the global strategies governing bottom-up processes. That is, top-down processes refer to the cognitive preplanning not only of narrative structure but also of strategies underlying the use of cohesive devices. Bottom-up processes, in turn, would need to be redefined as the linguistic expression of these organizational strategies.

Limitations of the study

Any discussion of my results is, of course, not complete without mentioning the limitations of the present study and desiderata for future research. Several limitations have already been mentioned in previous sections. However, several additional ones have not yet been discussed and these will be presented in the following.

First of all, my study uses data from a limited number of non-randomly selected participants, since it explores data from intact classes in one specific program. Even though this can hardly be avoided in a study of language acquisition in an educational setting and even less so in a pilot project such as the one my data was collected in, it also means that results may not be fully generalizable. Some subgroups are too small, for example, even to be included in a statistical analysis and can thus not be considered representative (the MB group, cf. ch. 4.1.2). At the same time the specific make-up of these classes with respect to the three experience groups may limit the comparability with other immersion programs. However, the results obtained for differences between groups and the development from first to fourth grade are so solid that these factors should not constitute a significant danger to reliability. An additional limitation, which is directly attributable to the number of participants, is that no counterbalancing could be conducted in the administration of

the task in order to avoid excessively small subgroups; reasons for the order chosen were discussed in ch. 4.1. Thus, it would be desirable to repeat this study with a larger number of (randomly-selected) participants from different programs so that all subgroups are big enough to be representative and the task order can be counterbalanced. At the same time it would be desirable, as discussed in a previous section, to collect as many learner variables with potentially predictive value as possible.

Secondly, the present study focused on a quantitative analysis of coherence and cohesion. However, as my results have shown, several areas which were not considered may need a closer investigation with a more qualitative approach in the future. Thus, a detailed analysis of the subcategories of cohesion and coherence would be desirable to discover more qualitative differences among participants. Are there differences in the linguistic means used to establish connective or referential ties, for example, which are attributable to grade and/or experience group? Are there differences as to what and how much information is included in the SETTING component or how the ENDING is expressed? Similarly, a more detailed look at the relationship between cohesion and coherence should be conducted, which takes a look at correlations between subcategories. Lexical and referential ties, for example, have been found significant predictors at least of coherence ratings (Bae 2001: 74).

Especially with respect to coherence an investigation of participants' L1 data analogous the present approach should be conducted in order to confirm whether differences between first graders with and without L2 preschool experience are really attributable to linguistic differences. With respect to cohesion an analysis of L1 data could point out whether participants make systematic differences in the use of cohesive devices and whether the general order of frequency found in the L2 English data is confirmed for participants' L1 German data. At the same time the approach to cohesion chosen in the present study should be complemented with a functional analysis, especially of connectives. Even more fundamentally, the use of cohesive devices in L2 discourse should be compared with L2 listeners' (perceived) need for cohesive devices in (narrative) discourse comprehension – an issue that is far from resolved in monolingual L1 research as well.

Further areas of research that future studies should explore are, for example, a comparison with monolingual L1 data, in relation to which previous studies on L2 learners have reported an over- as well as an underuse of cohesive devices (cf. Bae

2001, Viberg 2001), and a comparison with results achieved in traditional foreign language teaching. With respect to theory building it would be more than desirable to link existing research and theories on the "next lower" level of (L2) production, i.e. syntax, more strongly with discourse phenomena and to review existing (L2) acquisition theories critically in the light of (L2) discourse development.

Conclusions for the effectiveness of immersion programs

Discourse features are an important component of language proficiency and are therefore included in all major language tests and models of language proficiency (cf. ch.1). The results of the present study show that participants in the Kiel IM program benefit from IM education with respect to one of the most important genres of discourse, namely narratives. Make-believe stories, such as the ones used for data elicitation in the Kiel project, require a largely autonomous construction of text with a culturally pre-defined structure and they have thus been identified as one of the most challenging types of narrative discourse (cf. ch. 1 and ch. 2). The results presented in previous chapters showed that participants' L2 discourse proficiency improves significantly over the duration of an IM program such as the one under consideration, i.e. early partial IM at elementary school (for some participants in combination with IM in preschool, cf. ch. 4.1). While strong oral communication skills have been observed repeatedly for IM students (cf. ch. 1), especially in comparison with their peers in regular foreign language teaching programs, my results show that IM students also benefit with respect to academically valued discourse structuring skills – whether it be narrative structure or the linguistic connectedness of clauses and sentences. This beneficial effect is observed for both bilingual preschool (significant differences between experience groups in first grade) and subsequent immersion in elementary school (development from first to fourth grade independent of experience group).

The present study also contributes to two areas of immersion research, which have just recently been identified as "neglected" (Tucker & Dubiner 2008: 272): Newcomers to immersion programs and (knowledge) transfer. In the present study first graders without L2 preschool experience represent such a newcomer group. Results show that these children lag behind initially (first grade results) but that they are able to compensate this initial disadvantage if given sufficient time. Since only first and fourth grade results were studied, it is not entirely clear how much time is needed. My results show, however, that one year is not enough, while four years are

sufficient. Evidence from other studies conducted in the Kiel IM project, for example on verbal morphology, point to two years as the necessary time span to catch up (cf. Wode 2009: 37). Further testing is needed to see whether this also applies to IM programs with other set-ups, especially those with a lower percentage of instruction in the L2.

Independently of bilingual or monolingual preschool experience children thus benefit from attending an immersion program. Since any differences attributable to preschool experience have evened out until the end of fourth grade, does this mean that attending a bilingual preschool is superfluous? This conclusion would be premature. First of all, we do not know how a class without one or the other group would perform. It is not unlikely, however, that a class consisting entirely of children without L2 preschool experience would perform less strongly, since in heterogeneous classes inexperienced children can benefit from linguistically more advanced ones, especially in the initial phases.

Secondly, this study compares only first and fourth grade data. That is, results do not indicate how early the group with bilingual preschool experience arrives at the level it has reached by the end of fourth grade. First grade results suggest, however, that this may be earlier than for children without prior experience (at least for coherence). That is, the inexperienced group needs a comparatively longer exposure time to the L2 in elementary school in order to reach the same level of proficiency. This in turn has consequences for program planning, especially if an L3 is supposed to be introduced into the program. That is, an L3 could be introduced earlier for a group with L2 preschool experience than for an inexperienced group.

A more viable conclusion would therefore be that the significantly stronger performance of first graders with L2 preschool experience shows the beneficial effect of combining preschool and elementary school immersion and that an immersion program should indeed combine both educational levels in order to achieve optimal results.

With respect to knowledge transfer, coherence results point to a positive transfer of narrative schema, a language-independent cognitive structure. Earlier research on monolingual L1 speakers (cf. ch. 3) showed that in my participants' age range (mean age 6;8 and 9;8) a strong development of coherence is to be expected in L1 discourse. The results of my study showed that an equally strong development can be observed in L2 texts: A comparison of my results and those of previous

studies (cf. ch. 3) thus strongly suggests that the availability of a story schema for L2 discourse production means a general availability, i.e. the application of the schema is not bound to any particular language, even if its successful implementation can be hindered by linguistic difficulties. However, an analysis of L1 data would need to be conducted to fully confirm this claim.

Finally, the popular myth of cognitive deficits due to early bilingualism and/or schooling in an L2 has once more been shown to lack empirical evidence. That is, cognitive development – as evident from narrative coherence results – is not impeded by attending an immersion program. On the contrary, immersion programs, such as the one whose results were analyzed in the present study, are highly successful in fostering both cognitive and linguistic skills.[1]

[1] But cf. the previous section on a discussion of the construct validity of cohesion.

9 Bibliography

Aarssen, J., Akinçi, M.-A., Yağmur, K., 2001. Development of clause linkage in narratives: A comparison of Turkish children in Australia, France, the Netherlands and Turkey. In: Barreña, A., Ezeizabarrena, M.-J., Idiazabal, I. & MacWhinney, B. (eds.). *Research on child language acquisition: Proceedings of the 8th conference of the IASCL*. San Sebastián: University of the Basque Country, 41-56.

Akinçi, M.-A., Jisa, H. & Kern, S., 2001. Influence of L1 Turkish on L2 French narratives. In: Verhoeven, L. & Strömqvist, S. (eds.). *Narrative development in a multilingual context*. Amsterdam & Philadelphia, PA: John Benjamins, 188-208.

Aksu-Koç, A. & von Stutterheim, C., 1994. Temporal relations in narrative: Simultaneity. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 393-455.

Almgren, M., Beloki, L., Idiazabal, I. & Manterola, I., 2008. Acquisition of Basque in successive bilingualism: Data from oral storytelling. In: Siemund, P. & Kintana, N. (eds.). *Language contact and contact languages*. Amsterdam & Philadelphia: John Benjamins, 239-259.

Almgren, M., Idiazabal, I., Manterola, I. & Beloki, L., 2007. Language acquisition in Basque-Spanish successive bilinguals: Narrative skills in five year old children. Paper presented at the 6th International Symposium on Bilingualism. Hamburg, May 30- June 2, 2007.

Anderson, R.C. & Pearson, P.D., 1984. A schema theoretic view of basic processes in reading comprehension. In: Pearson, P.D. (ed.). *Handbook of reading research*. New York: Longman, 255-291.

Anderson, R.C., 1994 [1984]. Role of the reader's schema in comprehension, learning and memory. In: Ruddell, R.B., Ruddell, M.R. & Singer, H. (eds.). *Theoretical models and processing of reading.* 4th edition. Newark, DE: International Reading Association, 469-482.

Bachem, J., 2004. *Lesefähigkeiten deutscher Kinder im frühen englischen Immersionsunterricht*. Kiel University: M.A. thesis.

Bachmann, L.F., 1990. *Fundamental considerations in language testing*. Oxford, UK et al.: Oxford University Press.

Bae, J., 2001. Cohesion and coherence in children's written English: Immersion and English-only classes. *Issues in Applied Linguistics*, 12, 1, 51-88.

Baker, S. & MacIntyre, P.D., 2003. The role of gender and immersion in communication and second language orientations. *Language Learning*, 53, 1, 65-96.

Bamberg, M., 1986. Studies on morphological and syntactic development. *Linguistics*, 24, 227-284.

Bamberg, M., 1987. *The acquisition of narratives: Learning to use language*. Berlin et al.: Mouton de Gruyter.

Bamberg, M., 1994. Development of linguistic forms: German. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 189-238.

Bamberg, M. & Marchman, V., 1990. What holds a narrative together? The linguistic encoding of episode boundaries. *IprA Papers in Pragmatics*, 4, 1/2, 58-121.

Bamberg, M. & Marchman, V., 1991. Binding and unfolding: Towards the linguistic construction of narrative discourse. *Discourse Processes*, 14, 277-305.

Bamberg, M. & Marchmann, V., 1994. Foreshadowing and wrapping up in narrative. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 555-590.

Bamberg, M. & Moissinac, L., 2002. Discourse development. In: Graesser, A.C. Gernsbacher, M.A. & Goldman, S.R. (eds.). *The handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum Associates, 395-437.

Bartlett, F.C., 1932. *Remembering*. Cambridge, UK: Cambridge University Press.

Beigman Klebanov, B. & Shamir, E., 2005. Lexical cohesion: Some implications of an empirical study [online]. In: Bernadette Sharp (ed*.). Natural Language Understanding and Cognitive Science, Proceedings of the 2nd International Workshop on Natural Language Understanding and Cognitive Science, NLUCS 2005*. INSTICC Press, 13-21. Available at <URL: http://www.cs.huji.ac.il/~beata/#NLUCS>. Accessed July 5, 2006.

Bennet-Kastor, T., 1986. Cohesion and predication in child narrative. *Journal of Child Language,* 13, 353-370.

Berman, R.A., 1988. On the ability to relate events in narrative. *Discourse Processes*, 11, 469-497.

Berman, R.A., 1995. Narrative competence and storytelling performance: How children tell stories in different contexts. *Journal of Narrative and Life History*, 5, 4, 285-313.

Berman, R.A., 2001a. Narrative development in multilingual contexts: A cross-linguistic perspective. In: Verhoeven, L. & Strömqvist, S. (eds.). *Narrative development in a multilingual context*. Amsterdam & Philadelphia: John Benjamins, 419-428.

Berman, R.A., 2001b. Setting the narrative scene: How children begin to tell a story. In: Nelson, K., Aksu-Koç, A. & Johnson, C.E. (eds.). *Children's Language, Volume 10: Developing narrative and discourse competence*. Mahwah, NJ: Lawrence Erlbaum Associates, 1-30.

Berman, R.A. & Slobin, D.I. (eds.), 1994. *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berman, R.A. & D.I. Slobin, 1994a. IB: Goals and procedures. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 17-35.

Berman, R.A. & D.I. Slobin, 1994b. IIA: Narrative structure. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 39-84.

Berman, R.A. & D.I. Slobin, 1994c. IIIA: English. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 127-187.

Berman, R.A. & D.I. Slobin, 1994d. VA: Becoming a proficient speaker. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 597-610.

Biber, D., Conrad, S. & Leech, G., 2002. *Longman student grammar of spoken and written English*. Harlow, UK: Pearson Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E., 1999. *Longman grammar of spoken and written English*. Harlow, UK: Longman.

Bley-Vroman, R., 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning*, 33, 1, 1-17.

Bloor, T. & Bloor, M., 1995. *The functional analysis of English*. London, New York et al.: Arnold.

Botvin, G.J. & Sutton-Smith, B., 1977. The development of structural complexity in children's fantasy narratives. *Developmental Psychology*, 13, 4, 377-388.

Bransford, J.D. & Johnson, M.K. 1972. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726.

Brewer, W.F., 1985. The story schema: Universal and culture-specific properties. In: Olson, D.R., Torrance, N., Hildyard, A. (eds.). *Literacy, language and learning: The nature and consequences of reading and writing*. Cambridge, UK: Cambridge University Press, 167-94.

Bublitz, W., Lenk, U. & Ventola, E. (eds.), 1999. *Coherence in spoken and written discourse: How to create it and how to describe it*. Amsterdam & Philadelphia, PA: John Benjamins.

Burmeister, P. & Daniel, A. 2002 How effective is late partial immersion? Some findings from a secondary school program in Germany. In *An integrated view of language development. Papers in honor of Henning Wode*, P. Burmeister, T. Piske & A. Rohde (eds), 499-515. Trier: Wissenschaftlicher Verlag Trier.

Burmeister, P. & Pasternak, R., 2004. Früh und intensiv: Englische Immersion in der Grundschule am Beispiel der Claus-Rixen-Grundschule Altenholz, *Mitteilungsblatt FMF-Landesverband Schleswig-Holstein,* August 2004, 24-30.

Cain, K., 2003. Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology*, 21, 335-351.

Cameron, C.A., Lee, K., Webster, S., Munro, K., Hunt, A.K. & Linton, M.J., 1995. Text cohesion in children's narrative writing. *Applied Psycholinguistics*, 16, 257-269.

Chafe, W.L., 1980. *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Clark, H.H., 1994. Discourse in production. In: Gernsbacher, M.A. (ed.). *Handbook of psycholinguistics*. San Diego, CA: Academic Press, 985-1021.

The Commission of the European Communities, 2003. *Promoting Language Learning and Linguistic Diversity: An Action Plan 2004-2006*. Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions. COM (2003) 449 final, 24.07.2003.

Cook, G., 1989. *Discourse*. Oxford, UK et al.: Oxford University Press.

Cook, V., 1999. Going beyond the native speaker in language teaching, *TESOL Quarterly, 33*, 185-209.

Cook-Gumperz, J. & Green, J. L., 1986. A sense of story: Influences on children's storytelling ability. In: Tannen, D. (ed.). *Coherence in spoken and written discourse*. Norwood, NJ: Ablex, 201- 218.

Cortina, J.M., 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* 78, 1, 98-104.

The Council of Europe, 2001. *Common European Framework of reference for languages: Learning, Teaching, Assessment.* Cambridge, UK: Cambridge University Press.

Crowhurst, M., 1987. Cohesion in argument and narration at three grade levels. *Research in the Teaching of English*, 21, 2, 185-201.

Cruse, A., 2004. *Meaning in language: An introduction to semantics and pragmatics*. 2nd edition. Oxford, UK et al.: Oxford University Press.

Crystal, D., 2003. *A dictionary of linguistics and phonetics*. 5th edition. Malden, Mass. et al.: Blackwell.

Cutler, A. & Clifton, C. Jr., 2000. Comprehending spoken language: a blueprint of the listener. In: Brown, C.M. & Hagoort, P. (eds.). *The neurocognition of language*. Oxford, UK et al.: Oxford University Press, 123-166.

Cutting, J., 2002. *Pragmatics & Discourse: A resource book for students*. London, UK et al.: Routledge.

de Beaugrande, R.-A. & Dressler, W., 1981. *Introduction to text linguistics*. Harlow, UK: Longman.

Eckstein, P.P., 2008. *Angewandte Statistik mit SPSS: Praktische Einführung für Wirtschaftswissenschaftler*. 6th edition. Wiesbaden: Gabler.

Eikmeyer, H.-J., Kindt, W., Laubenstein, U., Lisken, S., Rieser, H. & Schade, U., 1995. Coherence regained. In: Rickheit, G. & Habel, C. (eds.). *Focus and*

coherence in discourse processing. Berlin & New York: Walter de Gruyter, 115-142.

Ellis, R., 1994. *The study of second language acquisition*. Oxford, UK et al.: Oxford University Press.

Eysenck, M.W., 2001. *Principles of cognitive psychology*. 2nd edition. Hove, UK: Psychology Press.

Field, A., 2009. *Discovering statistics using SPSS*. 3rd edition. Los Angeles, CA et al.: Sage.

Fitzgerald, J. & Spiegel, D.L., 1986. Textual cohesion and coherence in children's writing. *Research in the Teaching of English*, 20, 3, 263-280.

Fivush, R. & Slackman, E., 1986. The acquisition and development of scripts. In: Nelson, K. (ed.). *Event knowledge: Structure and function in development*. Hillsdale, NJ: Lawrence Erlbaum Associates, 71- 96.

Foltz, P.W., 2003. Quantitative cognitive models of text and discourse processing. In: Graesser, A.C., Gernsbacher, M.A. & Goldman, S.R. (eds.). *The handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum Associates, 487-523.

Foster-Cohen, S., 2001. First language acquisition … second language acquisition: 'What's Hecuba to him or he to Hecuba?'. *Second Language Research*, 17, 4, 329-344.

Genesee, F., 1987. *Learning through two languages*. Cambridge, Mass.: Newbury House.

Gernsbacher, M.A. & Givón, T. (eds.), 1995. *Coherence in spontaneous text*. Amsterdam & Philadelphia, PA: John Benjamins.

Gernsbacher, M.A. & Givón, T. (eds.), 1995a. Introduction: Coherence as a mental entity. In: Gernsbacher, M. A. & Givón, T. (eds.). *Coherence in spontaneous text*. Amsterdam & Philadelphia, PA: John Benjamins, vii-x.

Givón, T., 1995. Coherence in text vs. coherence in mind. In: Gernsbacher, M. A. & Givón, T. (eds.). *Coherence in spontaneous text*. Amsterdam & Philadelphia, PA: John Benjamins, 59-115.

Gombert, J.E., 1992. *Metalinguistic development*. Chicago, IL: University of Chicago Press.

Goswami, U., 1998. *Cognition in children*. Hove, UK: Psychology Press.

Graesser, A.C., Gernsbacher, M.A. & Goldman, S.R., 2003. Introduction to the handbook of discourse processes. In: Graesser, A.C., Gernsbacher, M.A. & Goldman, S.R. (eds.). *The handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum Associates, 1-23.

Greenbaum, S., 1996. *The Oxford English grammar*. Oxford, UK et al.: Oxford University Press.

Grice, H.P., 1975. Logic and conversation. In: Cole, P. & Morgan, J. (eds.). *Syntax and Semantics, Vol. 3: Speech Acts*. New York: Academic Press, 41-58.

Gumperz, J.J., Kaltman, H. & O'Connor, M.C., 1986. Cohesion in spoken and written discourse: Ethnic style and the transition to literacy. In: Tannen, D. (ed.). *Coherence in spoken and written discourse*. Norwood, NJ: Ablex, 3-19.

Halliday, M.A.K. & Hasan, R., 1976. *Cohesion in English*. London: Longman.

Halliday, M.A.K. & Hasan, Ruqaiya, 1997 [1985]. *Language, context and text: Aspects of language in a social-semiotic perspective*. Victoria: Deakin University Press.

Harley, B., Cummins, J., Swain, M. & Allen, P., 1990. The nature of language proficiency. In: Harley, B., Allen, P., Cummins, J. & Swain, M. (eds.), *The development of second language proficiency*. Cambridge, UK et al.: Cambridge University Press, 7-25.

Hasan, R., 1984. Coherence and cohesive harmony. In: Flood, J. (ed.), *Understanding reading comprehension: Cognition, language and the structure of prose*. Newark, Del.: International Reading Association, 181-220.

Hatch, E. & Lazaraton, A., 1991. *The research manual: Design and statistics for applied linguistics.* Boston, MA: Heinle & Heinle.

Hellmann, C., 1995. The notion of coherence in discourse. In: Rickheit, G. & Habel, C. (eds.). *Focus and coherence in discourse processing*. Berlin & New York: Walter de Gruyter, 190-202.

Hendrickson, V. & Shapiro,L.R., 2001. Cohesive reference devices. *Journal of Psychological Inquiry*, 6, 1, 17-22.

Herrman, C. & Fiebach, C., 2004. *Gehirn und Sprache*. Frankfurt/ Main: Fischer.

Hickmann, M., 2003. *Children's discourse: Person, space and time across languages*. Cambridge, UK: Cambridge University Press.

Hickmann, M., 1996. Discourse organization and the development of reference to person, space, and time. In: Fletcher, P. & MacWhinney, B. (eds.). *The handbook of child language*. Malden, Mass. et al.: Blackwell, 194-218.

Hickmann, M., 1991. The development of discourse cohesion: Some functional and cross-linguistic issues. In: Piéraut- Le Bonniec, G. & Dolitsky, M. (eds.). *Language bases... discourse bases: Some aspects of contemporary French-language psycholinguistics research*. Amsterdam: John Benjamins, 157-185.

Hicks, D., 1991. Kinds of narrative: Genre skills among first graders from two communities. In: McCabe, A. & Peterson, C. (eds.). *Developing narrative structure*. Hillsdale, NJ: Lawrence Erlbaum Associates, 55-87.

Hoey, M., 1996. *Patterns of lexis in text*. Oxford, UK et al.: Oxford University Press.

Huddleston, R. & Pullum, G.K., 2002. *The Cambridge grammar of the English language.* Cambridge, UK et al.: Cambridge University Press.

Hudson, J.A. & Shapiro, L.R., 1991. From knowing to telling: The development of children's scripts, stories, and personal narratives. In: McCabe, A. & Peterson, C. (eds.). *Developing narrative structure*. Hillsdale, NJ: Lawrence Erlbaum Associates, 89-136.

Hüttner, J. & Rieder-Bünemann, A., 2007. The effect of CLIL instruction on children's narrative competence. *VIEWS,* 16, 3 (CLIL Special Issue 2), 20-27.

Hunt, K.W., 1965. *Grammatical structures written at three grade levels*. Champaign, IL: National Council of Teachers of English.

Joanette, Y. & Brownell, H.H. (eds.), 1990. *Discourse ability and brain damage: Theoretical and empirical perspectives*. New York et al.: Springer.

Johnson-Laird, P.N., 1991. Mental models. In: Posner, M.I. (ed.). *Foundations of cognitive science*. Cambridge, MA: MIT Press, 469-499.

Johnson, R.K. & Swain, M. (eds.), 1997. *Immersion Education: International perspectives*. Cambridge, Mass.: Cambridge University Press.

Johnstone, B., 2002. *Discourse Analysis*. Malden, Mass & Oxford, UK: Blackwell.

Kamalski, J., Sanders, T. & Lentz, L., 2008. Coherence marking, prior knowledge, and comprehension of informative persuasive texts: Sorting things out. *Discourse Processes*, 45, 323- 345.

Kang, J.Y., 2005. Written narratives as an index of L2 competence in Korean EFL learners. *Journal of Second Language Writing*, 14, 4, 259-279.

Kang, J.Y., 2004. Telling a coherent story in a foreign language: Analysis of Korean EFL learners' referential strategies in oral discourse. *Journal of Pragmatics,* 36, 1976-1990.

Kang, J.Y., Kim, Y.-S. & Pan, B.A., 2009. Five-year-olds' book talk and story retelling: Contributions of mother-child joint bookreading. *First Language*, 29, 3, 243-265.

Karmiloff-Smith, A., 1985. Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes*, 1, 61-85.

Kayser, H., 1989. Some aspects of language understanding, language production, and intercomprehension in verbal interaction. In: Heydrich, W., Neubauer, F., Petöfi, J.S. & Sözer, E. (eds.). *Connexity and coherence: Analysis of text and discourse*. Berlin & New York: Walter de Gruyter, 342-365.

Kellogg, R.T., 1995. *Cognitive psychology*. Thousand Oaks, CA et al.: Sage.

Kersten, K., 2009. *Verbal inflections in L2 child narratives: A study of lexical aspect and grounding*. Trier: Wissenschaftlicher Verlag Trier.

Kersten, K., Imhoff, C. & Sauer, B., 2002. The Acquisition of English Verbs in an Elementary School Immersion Program in Germany. In: Burmeister, P., Piske, T. & Rohde, A. (eds.). *An Integrated View of Language Development: Papers in Honor of Henning Wode.* Trier: Wissenschaftlicher Verlag Trier, 473-497.

Khodadady, E. & Herriman, M., 2000. Schemata Theory and Selected Response Item Tests: From Theory to Practice. In: Kunnan, A.J. (ed.). *Fairness and validation in language assessment*. Cambridge, UK: Cambridge University Press, 201-222.

Kintsch, W., 1998. *Comprehension: A paradigm for cognition*. Cambridge, UK et al.: Cambridge University Press.

Kintsch, W., 1995. How readers construct situation models for stories: The role of syntactic cues and causal inferences. In: Gernsbacher, M.A. & Givón, T. (eds.). *Coherence in spontaneous text*. Amsterdam & Philadelphia, PA: John Benjamins, 139-160.

Kintsch, W., 1982. *Gedächtnis und Kognition*. Translated by Angelika Albert. Berlin et al.: Springer.

Kintsch, W., 1974. *The representation of meaning in memory*. New York et al.: John Wiley & Sons.

Kintsch, W. & van Dijk T. A., 1978. Toward a model of text comprehension and production. *Psychological Review*, 85, 5, 363-394.

Knott, A. & Sanders. T., 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30, 135-175.

Kreyß, J., 1995. Comprehension processes as a means for text generation. In: Rickheit, G. & Habel, C. (eds.). *Focus and coherence in discourse processing*. Berlin & New York: Walter de Gruyter, 143-169.

Kupersmitt, J. & Berman, R. A., 2001. Linguistic features of Spanish-Hebrew children's narratives. In: Verhoeven, L. & Strömqvist, S. (eds.). *Narrative development in a multilingual context*. Amsterdam & Philadelphia: John Benjamins, 277-317.

Labov, W., 1999 [1972]. The transformation of experience in narrative. In: Jaworski, A. & Coupland, N. (eds.). *The discourse reader*. London, UK & New York: Routledge, 221-235.

Labov, W. & Waletzky, J., 1967. Narrative analysis: Oral versions of personal experience. In: Helm, J. (ed.). *Essays on the verbal and visual arts*. Seattle: University of Washington Press, 12-44.

Lakshmanan, U. & Selinker, L., 2001. Analyzing interlanguage: How do we know what learners know? *Second Language Research*, 17, 4, 393-420.

Lanza, E., 2001. Temporality and language contact in narratives by children bilingual in Norwegian and English. In: Verhoeven, L. & Strömqvist, S. (eds.). *Narrative development in a multilingual context*. Amsterdam & Philadelphia: John Benjamins, 15-50.

Larsen-Freeman, D., 1991. Second Language Acquisition Research: Staking Out the Territory. *TESOL QUARTERLY*, 28, 2, 315-350.

Laurén, U., 1998. Narrative structures in the stories of immersion pupils in their second language. In: Arnau, J. & Artigal, J. (eds.). *Immersion programmes: A European perspective*. Barcelona: Universitat de Barcelona, 522- 531.

Lazaruk, W., 2007. Linguistic, academic, and cognitive benefits of French Immersion. *The Canadian Modern Language Review/ La Revue canadienne des languages vivantes*, 63, 5, 605-628.

Lenk, U., 1998. Discourse markers and global coherence in conversation. *Journal of Pragmatics*, 30, 245-257.

Levelt, W.J.M., 1999. Producing spoken language: A blueprint of the speaker. In: Brown, C. M. & Hagoort, P. (eds.). *The neurocognition of language*. Oxford, UK et al.: Oxford University Press, 83-122.

Löbner, S., 2002. *Understanding semantics*. London: Hodder Arnold.

Louwerse, M.M. 2004. A concise model of cohesion in text and coherence in comprehension. Memphis: Department of Psychology, 1-33, <http://madresearchlab.org/references.htm> (21 December 2005). English translation of: Un modelo conciso de cohesión en el texto y coherencia en la comprensión. *Revista Signos* 37: 41-58,

Lück, H.E. & Miller, R., 2005. *Illustrierte Geschichte der Psychologie*. Weinheim: Beltz.

Mackey, A. & Gass, S., 2005. *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mandel, R.G. & Johnson, N.S., 1984. A developmental analysis of story recall and comprehension in adulthood. *Journal of verbal learning and verbal behavior*, 23, 643-659.

Mandler, J.M. & Goodman, M.S., 1982. On the psychological validity of story structure. *Journal of verbal learning and verbal behaviour*, 21, 507-523.

Mandler, J.M. & Johnson, N.S., 1977. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111-151.

Manhardt, J. & Rescorla, L., 2002. Oral narrative skills of late talkers at ages 8 and 9. *Applied Psycholinguistics*, 23, 1-21.

Martin, J.R., 2001. Cohesion and Texture. In: Schiffrin, D., Tannen, D. & Hamilton, H. (eds.) *The handbook of discourse analysis*. Malden, Mass. et al.: Blackwell, 35-53.

Maschewski, C., 2002. *Pilotstudien zur Entwicklung kohäsiver Mittel im immersiven Englischunterricht von der 1. zur 2. Klasse der Claus-Rixen-Schule (Jahrgang 1999/2000)*. Kiel University: Unpublished diploma thesis (Staatsexamensarbeit).

Mayer, M., 1969. *Frog, where are you?*. New York: Puffin Pied Piper.

McCabe, A. & Peterson, C., 1991. *Developing narrative structure*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McCabe, A. & Peterson, C., 1991a. Getting the story: A longitudinal study of parental styles in eliciting narratives and developing narrative skills. In: McCabe, A. & Peterson, C. (eds.). *Developing narrative structure*. Hillsdale, NJ: Lawrence Erlbaum Associates, 217-253.

McCutchen, D. & Perfetti, C. A., 1982. Coherence and connectedness in the development of discourse production. *Text*, 2, 1-3, 113-139.

Michaels, S. & Collins, J., 1986. Oral discourse styles: Classroom interaction and the acquisition of literacy. In: Tannen, D. (ed.). *Coherence in spoken and written discourse*. Norwood, NJ: Ablex, 219-244.

Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig Holstein, 1997. *Lehrplan Grundschule* [online]. Available at <URL: http://lehrplan.lernnetz.de/intranet1/links/materials/1106920403.pdf>. Accessed September 1, 2009.

Minsky, M., 1975. A framework for representing knowledge. In: Winston, P.H. (ed.). *The psychology of computer vision*. New York: McGraw-Hill, 211-277.

Möller, C., 2009. The history and future of bilingual education: Immersion teaching in Germany and its Canadian origins. In: Rüdiger, P. & Gross, K. (eds.). *Translation of cultures*. Amsterdam & New York: Rodopi, 235-254.

Möller, C., 2009a. Messbarkeit und Entwicklung von Textlänge in mündlichen Lernerdaten am Beispiel der L2 Englisch. In: Lütge, C., Kollenrott, A.I., Ziegenmeyer, B. & Fellmann, G. (eds.). *Empirische Fremdsprachenforschung – Konzepte und Perspektiven*. Frankfurt a.M.: Peter Lang, 63-74.

Möller, C., 2003. *Erwerb und Entwicklung kohäsiver Bindungen im immersiven Englischunterricht*. Kiel University: Unpublished diploma thesis (Staatsexamensarbeit).

Montanari, S., 2004. The development of narrative competence in the L1 and L2 of Spanish-English bilingual children. *International Journal of Bilingualism*, 8, 4, 449-497.

Myles, F., 2003. The early development of L2 narratives: A longitudinal study. *Marges Linguistiques*, 5, 40-55.

Nelson, K. (ed.), 1986. *Event knowledge: Structure and function in development*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nelson, K., 1986a. Event knowledge and cognitive development. In: Nelson, K. (ed.). *Event knowledge: Structure and function in development*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1-19, 231-263.

Nelson, K. & Gruendel, J.-M., 1986. Children's scripts. In: Nelson, K. (ed.). *Event knowledge: Structure and function in development*. Hillsdale, NJ: Lawrence Erlbaum Associates, 21-46.

Neuner, J.L., 1987. Cohesive ties and chains in good and poor freshman essays. *Research in the Teaching of English*, 21, 1, 92-105.

Norment, N., 1995. Quantitative analysis of cohesive devices in Spanish and Spanish ESL in narrative and expository written texts. *Language Quarterly*, 33, 3-4, 135-159.

O'Neill, D.K. & Holmes, A. 2002. Young preschoolers' ability to reference story characters: The contribution of gesture and character speech. *First Language* 22: 73-103.

O'Neill, D.K., Pearce, M.J. & Pick, J.L., 2004. Preschool children's narratives and performance on the Peabody Individualized Achievement Test-Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language*, 24, 149-183.

Pan, B.A. & Snow, C.E., 1999. The development of conversational and discourse skills. In: Barrett, M. (eds.). *The development of language*. Hove, UK: Psychology Press, 229-249.

Park, E.S., 2004. The comparative fallacy in UG studies. In Purdy, J. (ed.). Commentaries on the Comparative Fallacy in Second Language Research [online]. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 4, 1. Available at <URL: http://journals.tc-library.org/index.php/tesol/issue/view/7>. Accessed February 5, 2010.

Pearson, P.D., 1982. A primer for schema theory. *The Volta Review*, 84, 25-33.

Perfetti, C.A., 1999. Comprehending written language: A blueprint of the reader. In: Brown, C.M. & Hagoort, P. (eds.). *The neurocognition of language*. Oxford, UK et al.: Oxford University Press, 167-208.

Peterson, C. & Dodsworth, P., 1991. A longitudinal analysis of young children's cohesion and noun specification in narratives. *Journal of Child Language*, 18, 397-415.

Peterson, C. & McCabe, A., 1991. Linking children's connective use and narrative macrostructure. In: McCabe, A. & Peterson, C. (eds.). *Developing narrative structure*. Hillsdale, NJ: Lawrence Erlbaum Associates, 29-53.

Peterson, C. & McCabe, A., 1983. *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York & London: Plenum.

Porte, G. K., 2002. *Appraising research in second language learning: A practical approach to critical analysis of quantitative research*. Amsterdam & Philadelphia: John Benjamins.

Pradl, G.M., 1979. Learning how to begin and end a story. *Language Arts*, 56, 1, 21-25.

Quasthoff, U.M. & Becker, T. (eds.), 2004. *Narrative interaction*. Amsterdam & Philadelphia: John Benjamins.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J., 1985. *A comprehensive grammar of the English language*. London et al.: Longman.

Reid, J., 1992. A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. *Journal of Second Language Writing*, 1, 2, 79-107.

Reilly, J.S., Bates, E.A., Marchman, V.A., 1998. Narrative discourse in children with early focal brain injury. *Brain and Language*, 61, 335-375.

Renkema, J., 2004. *Introduction to discourse studies*. Amsterdam & Philadelphia: John Benjamins.

Rickheit, G., Sichelschmidt, L. & Strohner, H., 1995. Economical principles in coherence management: A cognitive systems approach. In: Rickheit, G. & Habel, C. (eds.). *Focus and coherence in discourse processing*. Berlin & New York: Walter de Gruyter, 170-189.

Rieger, B., 1987. Wissensrepräsentation und empirische Semantik: Eine computerlinguistische Aufgabe. In: Pasternack, G. (eds.). *Theorie und Empirie*. Bremen: Bremen University Press, 99-149.

Rietveld, T. & van Hout, R., 1993. *Statistical techniques for the study of language and language behaviour.* Berlin & New York: Mouton de Gruyter.

Rohde, A., 2005. *Lexikalische Prinzipien im Erst- und Zweitspracherwerb*. Trier: Wissenschaftlicher Verlag Trier.

Rohde, A. & Tiefenthal, C., 2002. On lexical learning abilites. In: Burmeister, P., Piske, T. & Rohde, A. (eds.). *An Integrated View of Language Development: Papers in Honor of Henning Wode.* Trier: Wissenschaftlicher Verlag Trier, 449-471.

Rumelhart, D.E., 1980. Schemata: The building blocks of cognition. In: Spiro, R.J., Bruce, B. & Brewer, W.F. (eds.). *Theoretical issues in reading and comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates, 33-58.

Rumelhart, D.E., 1975. Notes on a schema for stories. In: Bobrow, D.C. & Collins. A. (eds.). *Representation and Understanding: Studies in Cognitive Science*. New York et al.: Academic Press, 211-236.

Sanford, A.J. & Moxey, L.M., 1995. Aspects of coherence in written language: A psychological perspective. In: Gernsbacher, M.A. & Givón, T. (eds.).

Coherence in spontaneous text. Amsterdam & Philadelphia, PA: John Benjamins, 161- 187.

Schank, R.C., 1975. The structure of episodes in memory. In: Bobrow, D.G. & Collins, A. *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 237- 272.

Schank, R.C. & Abelson, R., 1977. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schermer, F.J., 2006. *Lernen und Gedächtnis*. 4th edition. Stuttgart: Kohlhammer.

Schmitt, N., 1996. Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 4, 350-353.

Schneider, W. & Büttner, G., 2002. Entwicklung des Gedächtnisses bei Kindern und Jugendlichen. In: Oerter, R. & Montada, L. (eds.). *Entwicklungspsychologie*. 5th edition. Weinheim et al.: Beltz, 495-516.

Schubert, C., 2008. *Englische Textlinguistik: Eine Einführung*. Berlin: Erich Schmidt.

Scollon, R. & Scollon, S.B.K., 1986. Cooking it up and boiling it down: Abstracts in Athabaskan children's story retellings. In: Tannen, D. (ed.). *Coherence in spoken and written discourse*. Norwood, NJ: Ablex, 173- 197.

Segal, E.M. & Duchan, J.F., 1997. Interclausal connectives as indicators of structuring in narratives. In: Costermans, J. & Fayol, M. (eds.). *Processing interclausal relationships: Studies in the production and comprehension of text.* Mahwah, NJ: Lawrence Erlbaum Associates, 95-119.

Seidman, S., Nelson, K. & Gruendel, J., 1986. Make-believe scripts: The transformation of event representations in fantasy. In: Nelson, K. (ed.). *Event knowledge: Structure and function in development*. Hillsdale, NJ: Lawrence Erlbaum Associates, 161-187.

Severing, R. & Verhoeven, L., 2001. Bilingual narrative development in Papiamento and Dutch. In: Verhoeven, L. & Strömqvist, S. (eds.). *Narrative development in a multilingual context*. Amsterdam & Philadelphia: John Benjamins, 255-275.

Shapiro, L.R. & Hudson, J.A., 1997. Coherence and cohesion in children's stories. In: Costermans, J.& Fayol, M. (eds.). *Processing Interclausal Relationships: Studies in the production and comprehension of text*. Mahwah, NJ: Lawrence Erlbaum Associates, 23-48.

Shapiro, L.R. & Hudson, J.A., 1991. Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, 27, 960- 974.

Singer, M., 1990. *Psychology of language: An introduction to sentence and discourse processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Smit, U., 2008. The AILA Research Network - CLIL and immersion classrooms: Applied linguistics perspectives. *Language Teaching*, 41, 2, 295-298.

Smith, E.E., 1991. Concepts and induction. In: Posner, M.I. (ed.). *Foundations of cognitive science*. Cambridge, MA: MIT Press, 501-526.

Sodian, B., 2002. Entwicklung begrifflichen Wissens. In: Oerter, R. & Montada, L. (eds.). *Entwicklungspsychologie*. 5th edition. Weinheim et al.: Beltz, 443-468.

Solso, R.L., 2005. *Kognitive Psychologie*. Translated by Matthias Reiss. Berlin et al.: Springer.

Spiegel Online, 2008a. Merkel bedauert CDU-Beschluss zu deutscher Sprache [online]. Available at: <URL: http://www.spiegel.de/politik/deutschland/0,1518,594056,00.html>. Accessed December 3, 2008.

Spiegel Online, 2008b. Schwan wirft CDU Anti-Einwanderer-Wahlkampf vor [online]. Available at: <URL: http://www.spiegel.de/politik/deutschland/0,1518,594100,00.html>. Accessed December 3, 2008.

Spiegel, D.L. & Fitzgerald, J., 1990. Textual cohesion and coherence in children's writing revisited. *Research in the Teaching of English,* 24, 1, 48-66.

Stein, N.L., 1982. The definition of a story. *Journal of Pragmatics*, 6, 487-507.

Stein, N.L. & Albro, E.R., 1997. Building complexity and coherence: Children's use of goal-structured knowledge in telling stories. In: Bamberg, M (ed.). *Narrative development: Six approaches*. Mahwah, NJ: Lawrence Erlbaum Associates, 5-44.

Stein, N.L. & Glenn, C.G., 1979. An analysis of story comprehension in elementary children. In: Freedle, R.O. *New directions in discourse processing*. Norwood, NJ: Ablex, 53-120.

Stein, N.L. & Policastro, M., 1984. The concept of a story: A comparison between children's and teacher's viewpoints. In: Mandl, H., Stein, N.L. & Trabasso, T. (eds.). *Learning and comprehension of text*. Hillsdale, NJ: Lawrence Erlbaum Associates, 113-155.

Stenning, K. & Mitchell, L., 1985. Learning how to tell a good story: The development of content and language in children's telling of one tale. *Discourse Processes,* 8, 261-279.

Strömqvist, S. & Day, D., 1993. On the development of narrative structure in child L1 and adult L2 acquisition. *Applied Psycholinguistics*, 14, 135-158.

Tannen, D. (ed.), 1986. *Coherence in spoken and written discourse*. Norwood, NJ: Ablex.

Thorndyke, P.W., 1977. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology,* 9, 77-110.

Tiefenthal, C., 2009. *Fast mapping im natürlichen L2-Erwerb*. Trier: Wissenschaftlicher Verlag Trier.

Toolan, M.J., 1988. *Narrative: A critical linguistic introduction*. London & New York: Routledge.

Trabasso, T. & Rodkin, P.C., 1994. Knowledge of goal/plans: A conceptual basis for narrating 'Frog, where are you?'. In: Berman, R.A. & Slobin, D.I. (eds.). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, 85-106.

Trabasso, T., Suh, S. & Payton, P., 1995. Explanatory coherence in understanding and talking about events. In: Gernsbacher, M.A. & Givón, T. (eds.).

Coherence in spontaneous text. Amsterdam & Philadelphia, PA: John Benjamins, 189-214.

Tucker, R.G. & Dubiner, D., 2008. Concluding thoughts: Does the immersion pathway lead to multilingualism? In: Williams Fortune, T. & Tedick, D.J., *Pathways to multilingualism: Evolving perspectives on immersion education.* Clevedon, UK et al.: Multilingual Matters, 267-277.

Ungerer, F. & Schmid, H.-J., 2006. *An introduction to cognitive linguistics*. 2nd edition. Harlow, UK et al.: Pearson Education.

van de Velde, R., 1989. Man, verbal text, inferencing and coherence. In: Heydrich, W., Neubauer, F., Petöfi, J.S. & Sözer, E. (eds.). *Connexity and coherence: Analysis of text and discourse*. Berlin & New York: Walter de Gruyter, 174-214.

van Dijk, T.A., 1980. *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

van Dijk, T.A. & Kintsch, W., 1983. *Strategies of discourse comprehension.* New York: Academic Press.

Verhoeven, L., Aparici, M., Cahana-Amitay, D., van Hell, J., Kriz, S. & Viguié-Simon, A., 2002. Clause packaging in writing and speech: A cross-linguistic developmental analysis. In: Berman, R. A. & Verhoeven, L. (eds.). Cross-linguistic perspectives on the development of text-production abilities in speech and writing. *Written Language and Literacy* (special issue), 5, 2, 135-162.

Viberg, A., 2001. Age-related and L2-related features in bilingual narrative development in Sweden. In: Verhoeven, L. & Strömqvist, S. (eds.). *Narrative development in a multilingual context*. Amsterdam & Philadelphia: John Benjamins, 87-128.

von Berg, B., 2005. *Muttersprachliche Lesefähigkeiten bei L2-Immersionsuntericht: Eine Pilotstudie*. Kiel University: Unpublished diploma thesis (Staatsexamensarbeit).

Welt Online, 2008. CDU will Deutsch im Grundgesetz verankern [online]. Available at: <URL: http://www.welt.de/politik/article2814535/CDU-will-Deutsch-im-Grundgesetz-verankern.html>. Accessed December 3, 2008.

Westby, C.E., 1984. Development in narrative language abilities. In: Wallach, G.P. & Butler, K.G. (eds.). *Language learning disabilities in school-age children*. Baltimore, MD & London, UK: Williams & Wilkins.

Wigglesworth, G., 1990. Children's narrative acquisition: A study of some aspects of reference and anaphora. *First Language*, 10, 105-125.

Wode, H., 2009. *Frühes Fremdsprachenlernen in bilingualen Kindergärten und Grundschulen*. Braunschweig: Westermann.

Wode, H., 2004. *Frühes Fremdsprachenlernen: Englisch ab Kita und Grundschule: Warum? Wie? Was bringt es?* Kiel: Verein für frühe Mehrsprachigkeit an Kindertageseinrichtungen und Schulen (FMKS). Available at: <URL: http://www.fmks-online.de>. Accessed September 5, 2008.

Wode, H., 2000. Mehrsprachigkeit durch Kindergarten und Grundschulen: Chance oder Risiko? *Nouveaux Cahiers d'Allemand*, 19, 157-78.

Wode, H., 1998. A European perspective on immersion teaching: The German scenario. In: Arnau, J. & Artigal, J.M. (eds.). *Els programes d'immersió: Una perspectiva europea/ Immersion programmes: A European perspective*. Barcelona: Publicacions de la Universitat de Barcelona, 43-65.

Wode, H., 1995. Lernen in der Fremdsprache: *Grundzüge von Immersion und bilingualem Unterricht.* Ismaning: Hueber.

Wode, H., 1994. *Bilinguale Unterrichtserprobung in Schleswig-Holstein*. 2 volumes. Kiel: I&f Verlag.

Wode, H., Fischer, U., Pasternak, R.& Franzen, V., 2003. Frühes Englisch lernen im Altenholzer Verbund von Kita und Grundschule: Erfahrungen aus Praxis und Forschung zum Ende der 4. Klasse. In: Voss, B. & Stahlheber, E. (eds.). *Fremdsprachen auf dem Prüfstand: Innovation – Qualität – Evaluation*. Berlin: Pädagogischer Zeitschriftenverlag, 139-149.

Woods, A., Fletcher, P. & Hughes, A., 1986. *Statistics in language studies*. Cambridge, UK et al.: Cambridge University Press.

Yde, P. & Spoelders, M., 1985. Text cohesion: An exploratory study with beginning writers. *Applied Psycholinguistics*, 6, 407-416.

Zaunbauer, A.C.M., Bonerad, E.-M. & Möller, J., 2005. Muttersprachliches Leseverständnis immersiv unterrichteter Kinder. *Zeitschrift für Pädagogische Psychologie*, 19, 4, 263-65.

Zaunbauer, A.C.M. & Möller, J., 2007. Schulleistungen monolingual und immersiv unterrichteter Kinder am Ende des ersten Schuljahres. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 39, 3, 141-153.

10 Appendix

10.1 Individual cohesion scores ordered by frequency

Grade 1: Scores by frequency				
Overall cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
2,17	0,57	0	0	1,22
2,62	0,74	0	0	1,29
2,66	0,76	0,06	0	1,33
2,85	0,85	0,10	0	1,41
2,89	0,95	0,12	0	1,47
2,92	0,96	0,16	0	1,49
2,94	1,00	0,19	0	1,52
3,05	1,02	0,20	0	1,53
3,26	1,05	0,21	0,03	1,55
3,26	1,06	0,22	0,03	1,58
3,27	1,08	0,26	0,04	1,58
3,28	1,10	0,31	0,04	1,64
3,40	1,10	0,35	0,05	1,70
3,45	1,11	0,47	0,05	1,73
3,46	1,12	0,48	0,05	1,76
3,46	1,15	0,48	0,06	1,76
3,47	1,16	0,49	0,06	1,80
3,48	1,17	0,49	0,07	1,80
3,48	1,17	0,50	0,08	1,85
3,67	1,18	0,55	0,08	1,88
3,73	1,21	0,61	0,10	1,90
3,74	1,23	0,65	0,10	1,92
3,76	1,24	0,65	0,11	1,95
3,77	1,24	0,68	0,11	1,96
3,80	1,29	0,74	0,12	1,98
3,87	1,29	0,74	0,13	2,00
3,94	1,30	0,75	0,15	2,00
4,03	1,33	0,75	0,15	2,03
4,15	1,34	0,80	0,17	2,07
4,17	1,38	0,81	0,17	2,18
4,31	1,41	0,85	0,17	2,18
4,42	1,48	0,87	0,21	2,20
4,43	1,56	0,90	0,21	2,28
4,44	1,65	0,97	0,38	2,52
Total 34	**Total 34**	**Total 34**	**Total 34**	**Total 34**

Tab. 10.1 First graders' cohesive density scores in ascending order

Grade 1: Scores by frequency				
Total cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
3,43	0,76	0,46	0,03	1,58
3,49	0,96	0,47	0,05	1,68
3,58	1,04	0,48	0,05	1,69
3,59	1,07	0,48	0,07	1,69
3,63	1,07	0,49	0,08	1,73
3,64	1,10	0,49	0,08	1,76
3,67	1,10	0,50	0,09	1,80
3,69	1,13	0,51	0,10	1,80
3,72	1,13	0,53	0,10	1,81
3,74	1,15	0,53	0,12	1,83
3,74	1,16	0,53	0,13	1,87
3,75	1,18	0,53	0,13	1,88
3,76	1,20	0,55	0,14	1,88
3,78	1,21	0,56	0,14	1,89
3,88	1,22	0,57	0,14	1,89
3,88	1,23	0,58	0,15	1,89
3,90	1,24	0,58	0,15	1,93
3,90	1,24	0,59	0,16	1,94
3,91	1,26	0,64	0,17	1,94
3,95	1,29	0,64	0,17	1,95
3,97	1,30	0,65	0,17	1,99
3,98	1,31	0,66	0,17	2,00
3,98	1,32	0,67	0,18	2,00
4,02	1,32	0,70	0,18	2,00
4,04	1,34	0,72	0,19	2,00
4,07	1,36	0,73	0,19	2,02
4,13	1,36	0,76	0,19	2,02
4,21	1,37	0,76	0,19	2,02
4,34	1,37	0,79	0,20	2,21
4,44	1,38	0,79	0,20	2,22
4,64	1,49	0,86	0,21	2,25
4,65	1,62	0,91	0,23	2,31
Total 32	**Total 32**	**Total 32**	**Total 32**	**Total 32**

Tab. 10.2 Fourth graders' cohesive density scores in ascending order

10.2 Individual cohesion scores ordered by frequency and sex

Grade 1: Scores by frequency (male participants)				
Total cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
2,85	0,85	0,10	0	1,41
3,26	1,00	0,16	0	1,52
3,46	1,15	0,21	0	1,73
3,46	1,17	0,49	0	1,76
3,48	1,18	0,50	0,03	1,90
3,67	1,21	0,61	0,04	1,95
3,74	1,24	0,68	0,06	1,96
3,94	1,29	0,74	0,07	2,00
4,15	1,38	0,74	0,11	2,00
4,31	1,41	0,80	0,15	2,03
4,43	1,48	0,87	0,15	2,18
4,44	1,56	0,90	0,17	2,28
Total 12	**Total 12**	**Total 12**	**Total 12**	**Total 12**

Tab. 10.3 Male first graders' cohesive density scores in ascending order

Grade 1: Scores by frequency (female participants)				
Total cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
2,17	0,57	0	0	1,22
2,62	0,74	0	0	1,29
2,66	0,76	0,06	0	1,33
2,89	0,95	0,12	0	1,47
2,92	0,96	0,19	0,03	1,49
2,94	1,02	0,20	0,04	1,53
3,05	1,05	0,22	0,05	1,55
3,26	1,06	0,26	0,05	1,58
3,27	1,08	0,31	0,05	1,58
3,28	1,10	0,35	0,06	1,64
3,40	1,10	0,47	0,08	1,70
3,45	1,11	0,48	0,08	1,76
3,47	1,12	0,48	0,10	1,80
3,48	1,16	0,49	0,10	1,80
3,73	1,17	0,55	0,11	1,85
3,76	1,23	0,65	0,12	1,88
3,77	1,24	0,65	0,13	1,92
3,80	1,29	0,75	0,17	1,98
3,87	1,30	0,75	0,17	2,07
4,03	1,33	0,81	0,21	2,18
4,17	1,34	0,85	0,21	2,20
4,42	1,65	0,97	0,38	2,52
Total 22	**Total 22**	**Total 22**	**Total 22**	**Total 22**

Tab. 10.4 Female first graders' cohesive density scores in ascending order

Grade 4: Scores by frequency (male participants)					
Total cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Substitution density**	**Lexical density**
3,64	1,04	0,49	0,05		1,76
3,72	1,07	0,51	0,05		1,80
3,76	1,10	0,53	0,13		1,83
3,78	1,13	0,53	0,14		1,87
3,90	1,29	0,55	0,17		1,89
3,97	1,32	0,64	0,18		1,95
4,04	1,34	0,67	0,18		1,99
4,07	1,36	0,70	0,19		2,00
4,21	1,37	0,76	0,19		2,00
4,34	1,37	0,79	0,21		2,25
Total 10	**Total 10**	**Total 10**	**Total 10**		**Total 10**

Tab. 10.5 Male fourth graders' cohesive density scores in ascending order

Grade 4: Scores by frequency (female participants)					
Total cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Substitution density**	**Lexical density**
3,43	0,76	0,46	0,03		1,58
3,49	0,96	0,47	0,07		1,68
3,58	1,07	0,48	0,08		1,69
3,59	1,10	0,48	0,08		1,69
3,63	1,13	0,49	0,09		1,73
3,67	1,15	0,50	0,10		1,80
3,69	1,16	0,53	0,10		1,81
3,74	1,18	0,53	0,12		1,88
3,74	1,20	0,56	0,13		1,88
3,75	1,21	0,57	0,14		1,89
3,88	1,22	0,58	0,14		1,89
3,88	1,23	0,58	0,15		1,93
3,90	1,24	0,59	0,15		1,94
3,91	1,24	0,64	0,16		1,94
3,95	1,26	0,65	0,17		2,00
3,98	1,30	0,66	0,17		2,00
3,98	1,31	0,72	0,17		2,02
4,02	1,32	0,73	0,19		2,02
4,13	1,36	0,76	0,19		2,02
4,44	1,38	0,79	0,20		2,21
4,64	1,49	0,86	0,20		2,22
4,65	1,62	0,91	0,23		2,31
Total 22	**Total 22**	**Total 22**	**Total 22**		**Total 22**

Tab. 10.6 Female fourth graders' cohesive density scores in ascending order

10.3 Individual cohesion scores ordered by frequency and experience group

Grade 1: Scores by frequency (bili participants)				
Overall cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
2,66	0,85	0,06	0	1,41
2,85	0,96	0,16	0	1,49
3,26	1,00	0,21	0	1,52
3,27	1,02	0,31	0,04	1,55
3,28	1,06	0,35	0,05	1,58
3,40	1,08	0,47	0,05	1,64
3,45	1,10	0,49	0,06	1,70
3,46	1,10	0,49	0,06	1,73
3,46	1,12	0,50	0,07	1,76
3,67	1,15	0,55	0,08	1,88
3,73	1,17	0,61	0,10	1,92
3,74	1,17	0,65	0,11	1,95
3,76	1,18	0,65	0,12	1,96
3,77	1,21	0,68	0,13	1,98
3,87	1,23	0,74	0,15	2,00
3,94	1,24	0,74	0,15	2,03
4,03	1,29	0,75	0,17	2,07
4,15	1,34	0,75	0,17	2,20
4,44	1,56	0,87	0,21	2,28
Total 19	**Total 19**	**Total 19**	**Total 19**	**Total 19**

Tab. 10.7 Bili first graders' cohesive density scores in ascending order

Grade 1: Scores by frequency (mono participants)				
Overall cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
2,17	0,57	0	0	1,22
2,62	0,74	0	0	1,29
2,89	0,76	0,10	0	1,33
2,92	0,95	0,12	0	1,47
2,94	1,11	0,19	0	1,53
3,05	1,16	0,20	0,03	1,58
3,26	1,24	0,22	0,05	1,76
3,47	1,29	0,26	0,08	1,80
3,48	1,30	0,48	0,11	1,85
3,48	1,33	0,48	0,17	1,90
4,17	1,48	0,81	0,21	2,18
4,42	1,65	0,97	0,38	2,52
Total 12	**Total 12**	**Total 12**	**Total 12**	**Total 12**

Tab. 10.8 Mono first graders' cohesive density scores in ascending order

Grade 4: Scores by frequency (bili participants)				
Overall cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
3,43	0,76	0,47	0,03	1,58
3,49	1,04	0,48	0,08	1,69
3,67	1,07	0,49	0,09	1,80
3,69	1,10	0,50	0,10	1,87
3,72	1,13	0,53	0,13	1,88
3,74	1,13	0,56	0,14	1,89
3,75	1,15	0,59	0,14	1,89
3,88	1,18	0,64	0,15	1,93
3,90	1,20	0,64	0,17	1,94
3,97	1,23	0,65	0,17	2,00
3,98	1,31	0,66	0,18	2,00
4,02	1,32	0,72	0,18	2,02
4,21	1,32	0,73	0,19	2,02
4,34	1,36	0,76	0,19	2,21
4,44	1,36	0,79	0,21	2,22
4,64	1,49	0,91	0,23	2,25
Total 16	**Total 16**	**Total 16**	**Total 16**	**Total 16**

Tab. 10.9 Bili fourth graders' cohesive density scores in ascending order

Grade 4: Scores by frequency (mono participants)				
Overall cohesive density	**Reference density**	**Connective density**	**Ellipsis density**	**Lexical density**
3,58	0,96	0,46	0,05	1,68
3,59	1,07	0,49	0,08	1,69
3,63	1,10	0,51	0,10	1,73
3,64	1,16	0,53	0,12	1,76
3,76	1,21	0,57	0,14	1,80
3,78	1,22	0,58	0,15	1,81
3,88	1,24	0,58	0,16	1,83
3,91	1,24	0,67	0,17	1,88
3,95	1,26	0,70	0,17	1,94
3,98	1,29	0,76	0,19	2,00
4,13	1,38	0,79	0,20	2,02
4,65	1,62	0,86	0,20	2,31
Total 12	**Total 12**	**Total 12**	**Total 12**	**Total 12**

Tab. 10.10 Mono fourth graders' cohesive density scores in ascending order

10.4 Coherence and cohesion scores

10.4.1 Coherence scores: Overall measures

Participant	Total number of narrative components	Narrative index score
G1-C1-1[1]	6	5
G1-C1-2	6	7
G1-C1-3	1	2
G1-C1-4	5	7
G1-C1-5	6	7
G1-C1-6	8	9
G1-C1-7	2	4
G1-C1-8	3	6
G1-C1-10	6	9
G1-C1-11	1	2
G1-C1-13	2	4
G1-C1-14	7	7
G1-C1-15	2	4
G1-C1-16	10	9
G1-C1-17	6	7
G1-C1-18	4	6
G1-C8-1	6	7
G1-C8-2	5	7
G1-C8-3	10	9
G1-C8-4	3	6
G1-C8-5	7	9
G1-C8-6	2	4
G1-C8-7	1	2
G1-C8-8	1	2
G1-C8-9	1	2
G1-C8-10	1	2
G1-C8-11	3	4
G1-C8-13	9	7
G1-C8-14	7	9
G1-C8-15	3	6
G1-C8-16	1	2
G1-C8-18	2	4
G1-C8-20	8	9
G1-C8-21	10	9
G4-C1-1	9	9
G4-C1-2	8	9
G4-C1-3	7	9
G4-C1-4	10	9
G4-C1-5	11	9
G4-C1-6	11	9
G4-C1-7	11	9
G4-C1-8	10	9

[1] References to individual participants are coded as follows: Cohort-grade-participant number. 'C1-G1-1', for example, refers to child 1 from cohort 1 in grade 1, 'C1-G4-1' to the same child in grade 4 (longitudinal data set).

Participant	Total number of narrative components	Narrative index score
G4-C1-10	9	9
G4-C1-11	8	9
G4-C1-13	12	9
G4-C1-14	9	9
G4-C1-15	8	7
G4-C1-16	12	9
G4-C1-17	10	9
G4-C1-18	11	9
G4-C5-3	12	9
G4-C5-6	11	9
G4-C5-7	9	9
G4-C5-8	10	9
G4-C5-9	12	9
G4-C5-10	10	9
G4-C5-11	10	9
G4-C5-12	10	9
G4-C5-14	11	9
G4-C5-15	12	9
G4-C5-16	11	9
G4-C5-17	7	9
G4-C5-18	10	9
G4-C5-19	10	9
G4-C5-20	6	8
G4-C5-21	12	9

10.4.2 Coherence scores: Individual components

Participant	SETTING	INITIATING EVENT	SIMPLE REACTION	GOAL 1	GOAL 2	ATTEMPT 1	ATTEMPT 2	ATTEMPT 3	ATTEMPT 4	ATTEMPT 5	ATTEMPT 6	ATTEMPT 7	CONSEQUENCE	ENDING
G1-C1-1	0[1]	1	1	0	1	1	0	1	1	0	0	0	0	0
G1-C1-2	1	1	0	0	1	1	0	1	1	0	0	0	0	0
G1-C1-3	0	0	0	0	0	0	0	0	0	0	0	0	1	0
G1-C1-4	1	1	0	0	1	0	1	1	0	0	0	0	0	0
G1-C1-5	1	1	0	0	1	0	0	1	1	0	1	0	0	0
G1-C1-6	1	1	1	0	1	1	1	1	0	0	0	0	1	0
G1-C1-7	1	1	0	0	0	0	0	0	0	0	0	0	0	0
G1-C1-8	1	1	0	0	0	1	0	0	0	0	0	0	0	0
G1-C1-10	1	1	0	0	1	0	1	0	0	1	0	0	1	0
G1-C1-11	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G1-C1-13	1	1	0	0	0	0	0	0	0	0	0	0	0	0
G1-C1-14	1	1	0	1	1	0	0	0	1	1	0	0	0	1
G1-C1-15	0	1	0	0	0	1	0	0	0	0	0	0	0	0
G1-C1-16	1	1	0	1	1	0	1	1	1	0	1	0	1	1
G1-C1-17	0	1	0	1	1	0	1	1	0	0	0	0	0	1
G1-C1-18	1	1	0	0	0	0	0	0	0	0	0	0	1	1
G1-C8-1	1	1	0	0	1	1	1	1	0	0	0	0	0	0
G1-C8-2	1	1	0	0	1	1	0	1	0	0	0	0	0	0
G1-C8-3	1	1	0	1	1	1	1	1	0	0	1	0	1	1
G1-C8-4	1	1	0	0	0	0	0	0	0	0	0	0	1	0
G1-C8-5	1	1	0	0	1	0	1	1	0	0	1	0	1	0
G1-C8-6	1	0	0	0	0	0	0	0	0	0	0	0	1	0
G1-C8-7	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G1-C8-8	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G1-C8-9	0	1	0	0	0	0	0	0	0	0	0	0	0	0
G1-C8-10	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G1-C8-11	1	1	1	0	0	0	0	0	0	0	0	0	0	0
G1-C8-13	1	1	0	0	1	1	1	1	1	1	1	0	0	0
G1-C8-14	1	1	0	0	1	1	0	1	0	1	0	0	1	0
G1-C8-15	1	1	0	0	0	0	0	0	0	0	0	0	0	1
G1-C8-16	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G1-C8-18	0	1	0	0	0	0	0	0	0	0	0	0	1	0
G1-C8-20	1	1	1	0	1	1	0	1	0	0	1	0	1	0
G1-C8-21	1	1	0	0	1	1	0	1	1	1	1	1	1	0
G4-C1-1	1	1	0	1	1	1	1	1	0	0	0	0	1	1
G4-C1-2	1	1	0	1	1	1	0	1	0	0	0	0	1	1
G4-C1-3	1	1	0	1	1	1	0	1	0	0	0	0	0	1
G4-C1-4	1	1	0	1	1	1	1	1	0	1	0	0	1	1
G4-C1-5	1	1	1	1	1	1	0	1	0	1	1	0	1	1
G4-C1-6	1	1	1	1	1	1	1	1	0	0	1	0	1	1
G4-C1-7	1	1	0	1	1	1	1	1	0	1	1	0	1	1

[1] In all coherence analysis tables '0' stands for "component/ feature absent/ not realized", while '1' analogously stands for "component/ feature present/ realized."

Participant	SETTING	INITIATING EVENT	SIMPLE REACTION	GOAL 1	GOAL 2	ATTEMPT 1	ATTEMPT 2	ATTEMPT 3	ATTEMPT 4	ATTEMPT 5	ATTEMPT 6	ATTEMPT 7	CONSEQUENCE	ENDING
G4-C1-8	1	1	1	0	1	1	1	1	1	1	0	0	1	0
G4-C1-10	1	1	0	1	1	1	1	1	0	0	0	0	1	1
G4-C1-11	1	1	0	1	1	1	1	1	0	0	0	0	0	1
G4-C1-13	1	1	1	1	1	1	1	1	1	0	1	0	1	1
G4-C1-14	1	1	0	1	1	1	1	1	1	0	0	0	0	1
G4-C1-15	0	1	0	1	1	1	0	1	0	0	1	0	1	1
G4-C1-16	1	1	0	1	1	1	1	1	1	1	1	0	1	1
G4-C1-17	1	1	0	1	1	1	1	1	1	0	0	0	1	1
G4-C1-18	1	1	0	1	1	1	1	1	0	1	1	0	1	1
G4-C5-3	1	1	1	1	1	1	1	1	1	0	1	0	1	1
G4-C5-6	1	1	1	1	1	1	1	1	0	0	1	0	1	1
G4-C5-7	1	1	0	1	1	1	1	1	0	0	1	0	0	1
G4-C5-8	1	1	1	1	1	1	1	1	0	0	1	0	0	1
G4-C5-9	1	1	0	1	1	1	1	1	1	1	1	0	1	1
G4-C5-10	1	1	0	1	1	1	1	1	1	0	0	1	0	1
G4-C5-11	1	1	0	1	1	1	1	1	1	1	0	0	0	1
G4-C5-12	1	1	1	0	1	1	1	1	1	1	0	0	1	0
G4-C5-14	1	1	1	1	1	1	1	1	0	1	1	0	0	1
G4-C5-15	1	1	1	1	1	1	1	1	1	1	0	0	1	1
G4-C5-16	1	1	1	1	1	1	1	1	0	0	1	0	1	1
G4-C5-17	1	1	0	1	1	1	0	1	0	0	0	0	0	1
G4-C5-18	1	1	0	1	1	1	1	1	0	0	1	0	1	1
G4-C5-19	1	1	1	1	1	1	0	1	0	0	1	0	1	1
G4-C5-20	1	1	1	1	0	1	0	0	0	0	0	0	0	1
G4-C5-21	1	1	1	1	1	1	1	1	1	1	1	0	0	1

10.4.3 Cohesion scores

Participant	Total cohesive density	Reference density	Connective density	Substitution density	Ellipsis density	Lexical density
G1-C1-1	3,74	1,00	0,74	0,00	0,00	2,00
G1-C1-2	4,44	1,56	0,61	0,00	0,00	2,28
G1-C1-3	3,26	1,11	0,19	0,00	0,11	1,85
G1-C1-4	2,17	0,74	0,22	0,00	0,00	1,22
G1-C1-5	3,80	1,05	0,85	0,00	0,10	1,80
G1-C1-6	3,40	1,12	0,65	0,00	0,05	1,58
G1-C1-7	3,48	1,16	0,48	0,00	0,08	1,76
G1-C1-8	4,42	1,24	0,97	0,00	0,03	2,18
G1-C1-10	2,89	0,95	0,26	0,00	0,21	1,47
G1-C1-11	4,43	1,41	0,80	0,00	0,04	2,18
G1-C1-13	3,67	1,21	0,50	0,00	0,00	1,96
G1-C1-14	3,87	1,10	0,47	0,00	0,10	2,20
G1-C1-15	3,73	1,08	0,65	0,00	0,12	1,88
G1-C1-16	4,03	1,23	0,75	0,00	0,08	1,98
G1-C1-17	3,27	0,96	0,35	0,00	0,04	1,92
G1-C1-18	4,31	1,38	0,90	0,00	0,03	2,00
G1-C8-1	3,94	1,15	0,87	0,00	0,17	1,76
G1-C8-2	3,05	0,57	0,81	0,00	0,38	1,29
G1-C8-3	3,46	1,24	0,16	0,00	0,11	1,95
G1-C8-4	3,77	1,34	0,55	0,00	0,17	1,70
G1-C8-5	3,46	1,17	0,49	0,00	0,07	1,73
G1-C8-6	3,47	1,30	0,20	0,00	0,17	1,80
G1-C8-7	3,48	1,48	0,10	0,00	0,00	1,90
G1-C8-8	2,92	1,33	0,00	0,00	0,00	1,58
G1-C8-9	3,45	1,10	0,75	0,00	0,05	1,55
G1-C8-10	2,94	1,29	0,12	0,00	0,00	1,53
G1-C8-11	4,15	1,29	0,68	0,00	0,15	2,03
G1-C8-13	3,28	1,02	0,49	0,00	0,13	1,64
G1-C8-14	3,76	1,17	0,31	0,00	0,21	2,07
G1-C8-15	2,85	1,18	0,21	0,00	0,06	1,41
G1-C8-16	4,17	1,65	0,00	0,00	0,00	2,52
G1-C8-18	2,62	0,76	0,48	0,00	0,05	1,33
G1-C8-20	3,26	0,85	0,74	0,00	0,15	1,52
G1-C8-21	2,66	1,06	0,06	0,00	0,06	1,49
G4-C1-1	4,34	1,32	0,64	0,00	0,14	2,25
G4-C1-2	4,21	1,36	0,79	0,00	0,18	1,89
G4-C1-3	3,58	1,16	0,46	0,00	0,08	1,88
G4-C1-4	3,98	1,24	0,76	0,00	0,17	1,81
G4-C1-5	3,74	1,30	0,48	0,00	0,07	1,89
G4-C1-6	3,67	1,15	0,66	0,00	0,17	1,69
G4-C1-7	3,88	1,21	0,79	0,00	0,20	1,68
G4-C1-8	3,91	1,22	0,86	0,00	0,10	1,73
G4-C1-10	3,63	1,24	0,58	0,00	0,12	1,69
G4-C1-11	4,07	1,34	0,53	0,00	0,19	2,00
G4-C1-13	3,97	1,13	0,76	0,00	0,21	1,87
G4-C1-14	3,88	1,10	0,65	0,00	0,10	2,02
G4-C1-15	4,02	1,32	0,56	0,00	0,14	2,00

Participant	Total cohesive density	Reference density	Connective density	Substitution density	Ellipsis density	Lexical density
G4-C1-16	3,74	1,07	0,59	0,00	0,19	1,89
G4-C1-17	3,49	0,76	0,72	0,00	0,08	1,93
G4-C1-18	3,90	1,37	0,53	0,00	0,05	1,95
G4-C5-3	3,72	1,04	0,49	0,01	0,18	2,00
G4-C5-6	3,43	1,18	0,48	0,02	0,17	1,58
G4-C5-7	3,98	1,13	0,64	0,00	0,19	2,02
G4-C5-8	4,44	1,36	0,73	0,00	0,13	2,22
G4-C5-9	4,13	1,38	0,58	0,00	0,14	2,02
G4-C5-10	3,76	1,07	0,67	0,00	0,19	1,83
G4-C5-11	3,59	0,96	0,49	0,00	0,20	1,94
G4-C5-12	4,65	1,62	0,57	0,00	0,15	2,31
G4-C5-14	3,90	1,23	0,50	0,00	0,23	1,94
G4-C5-15	3,75	1,31	0,47	0,00	0,09	1,88
G4-C5-16	4,04	1,37	0,55	0,00	0,13	1,99
G4-C5-17	3,95	1,26	0,53	0,00	0,16	2,00
G4-C5-18	4,64	1,49	0,91	0,00	0,03	2,21
G4-C5-19	3,64	1,29	0,51	0,02	0,05	1,76
G4-C5-20	3,69	1,20	0,53	0,02	0,15	1,80
G4-C5-21	3,78	1,10	0,70	0,00	0,17	1,80