Can language play promote morphosyntactic learning?

A study about the impact of rhyme, melody and input structuring on gender-like category acquisition.

Inaugural-Dissertation

zur

Erlangung des Grades eines Doktors der Philosophie

in der

Fakultät für Philologie

der

RUHR-UNIVERSITÄT BOCHUM

vorgelegt

von

Johanna Bebout

Oktober 2013

Gedruckt mit der Genehmigung der Fakultät für Philologie der Ruhr-Universität Bochum

Referentin: Prof. Dr. Eva Belke

Korreferent: Prof. Dr. Ralf Klabunde

Tag der mündlichen Prüfung: 7. Mai 2014

Acknowledgements

This thesis would not have been possible without the professional and personal support of a great number of people.

First of all, I am endebted to my supervisors, Eva Belke and Ralf Klabunde. Eva Belke supported my personal and professional development in many ways. This dissertation would not have been possible without her consistent advice, encouragement and professional support. Ralf Klabunde was always very open-minded about my interdisciplinary research and supported my ideas from the start.

Next, I would like to thank Kathryn Bock for her guidance and support during my experimental investigation at the University of Illinois at Urbana-Champaign. Thank you for the access to your laboratory where I ran Experiment 3 of the present thesis and for giving me the opportunity to participate in current research discussions. I greatly appreciate the help of Laurel Brehm and Jonathan Scott Hajek from the University of Illinois at Urbana-Champaign for advice and guidance during my work at the laboratory. Also, I would like to thank Pienie Zwitserlood and her team from the University of Münster, for their professional support and for providing access to their laboratory where I ran parts of Experiments 1 and 2. I am indebted to Linda Kauffeldt for lending me her beautiful voice for the experiments and for composing the musical stimuli. Furthermore, I would like to show my gratitude to Claudia Kirschke for creating the lovely paintings I used in my experiments. The present experiments would not have been possible without the creative ideas and great effort of both of you.

The data collection would not have been possible without many participants in Germany and the USA, who volunteered in my experiments. I would like to thank Joshua Bebout, Maria Hosfeld and Martin Kitzinger for proofreading the present thesis and providing useful comments on my research. I also would like to thank Christina Clasmeier, Hendrike Frieg, Inga Hennecke, Anna Stielow and Friederike von Lehmden for their professional and personal support during the last four years.

I am greatful to my parents, my wonderful sister Linda, my amazing friend Katharina and my beloved husband Joshua for always believing in me and supporting me in every possible way.

Contents

1 Introduction

There is no doubt that our ability to accurately comprehend and produce sentences is crucially based on the ability to generalize morphosyntactic regularities. Grammar learning, such as learning about morphosyntactic forms, structures, regularities and paradigms can thus be considered as an essential part of language acquisition.

In many languages, the acquisition of gender and case plays a pivotal role in morphosyntactic learning. The acquisition of a gender-case system requires the learner to a) acquire the correct gender category for a noun and to b) acquire the case paradigm associated with each gender category. It is well-known that the German gender-case system is of high complexity and extremely hard to acquire, at least for certain learner groups. Many children growing up in Germany have severe difficulties with the acquisition of this system, especially children who learn German as a second language or those who are not exposed to frequent and rich German input in their early childhood. A similar pattern can be seen in children who receive sufficient input but have difficulties with the processing of the input, such as hearing impaired children. The difficulties are typically apparent by the wrong use of gender- and case-marked forms, such as determiners or the inflection of prenominal adjectives.

When considering the acquisition of both the German gender categories and the German case paradigm, an obvious picture emerges. Given that the German case paradigm can be described using clear regularities, these case regularities are typically efficiently acquired and accurately used from an early age on. When the gender category of a German noun is known, children with German as their second language tend to produce case markers more correctly (Jeuk, 2007). In contrast to the rule-driven German case paradigm, distinct regularities are lacking regarding the assignment of gender categories to many German monomorphemic nouns (e.g. Maratsos, 1983). It is thus not surprising that gender acquisition and assignment constitutes a major challenge for children who acquire German as their second language (e.g. Jeuk, 2007; Rösch, 2003).

Given that every fourth child in Germany acquires German as a second language, the question of how to support these children in their language acquisition process is of crucial relevance. What learning mechanisms enable

and facilitate a systematic language promotion of complex and opaque grammatical subclasses, such as the German gender categories?

Previous research has demonstrated that the acquisition of artificial gender-like subclasses is possible when nouns are phonologically marked (e.g. Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Frigo & MacDonald, 1998). However, even though there are a number of phonological noun cues indicating the gender subclasses German monomorphemic nouns belong to, virtually all of these rules have many exceptions. Additionally, these noun cues seem to be invalid for a great number of early acquired nouns, rendering their role in early learning of German questionable. Hence, some authors have argued that monomorphemic German nouns are largely arbitrarily assigned to gender subclasses (Maratsos, 1983 but see Köpcke, 1982). According to this view, a noun's syntactic environment, e.g. the association of different inflectional morphemes accompanying a noun, supplies the most significant cue towards its gender. Learners are able to exploit these syntactic cues, provided that they are presented by means of a systematic, well-structured input (Taraban, 2004). In everyday life, such learner-friendly input organization can for instance be found in forms of language play, such as nursery rhymes or childrens' songs.

Language play naturally connects a well-structured input with a rhyming or melodic structure or both. Prior research has demonstrated the potential value of music (e.g. Francois & Schön, 2010; Schön et al., 2008) and rhyming abilities (e.g. Wood, 2006) for language processing and language acquisition. Indeed, language play is proposed to be an excellent starting point for language learning and language promotion (e.g. E. Belke & Belke, 2006; Cook, 1997, 2000; Haueis, 1985). The present study was conducted in order to verify this claim empirically.

The purpose of the present thesis is to examine what input presentation helps support gender-like category induction when noun cues are missing. More precisely, I aimed to investigate the impact of three forms of input presentation on the acquisition of gender-like subclasses in an artificial language learning paradigm, namely structured presentation, rhyming presentation and melodic presentation. These presentation forms are combined in many forms of language play.

As to the investigation of the presentation forms, I was particularly interested in finding out how the exposure to linguistic input can be optimized in such a way that the assignment of phonologically unmarked nouns to gender-like

subclasses can be facilitated. Thus, participants were trained either with a well-structured, systematic input (Experiment 1, parts of Experiment 3) or with a random, unstructured input (Experiment 2, parts of Experiment 3). Furthermore, the present study investigated to what extent adding a rhyming structure, a melodic structure, or both to the linguistic input affects the acquisition of the artificial gender-like subclasses. Based on language play as one way to systematically promote grammar skills in preschool and school children, the present study focused on implicit ways of facilitating the acquisition of gender-like subclasses.

In chapter 2 of the present thesis, I will first review the evidence concerning the role of noun cues for gender assignment. Phonological and semantic noun cues, such as certain suffixes or semantically driven regularities, have been postulated to be critical for acquiring the correct gender category of a noun. Empirical research working with artificial or natural gender(-like) categories has confirmed this claim. However, as previously noted, German monomorpehmic nouns possess only few phonological or semantic noun cues as to their gender. I will briefly summarize the major assumptions and findings regarding the existence of German noun cues and their role in language learning, demonstrating that the reliability of phonological noun cues in German is questionable. Given this, there may be other factors which support the acquisition of gender assignment, at least for languages that lack powerful noun cues. Therefore, the second part of chapter 2 considers the potential value of syntactic cues for the acquisition of gender categories when phonological cues are absent. A brief summary of the theoretical background as well as empirical evidence for the utility of syntactic cues will be reviewed in this section. More specifically, I will introduce a study conducted by Taraban (2004), the results of which suggest that syntactic cues are sufficient to acquire artificial gender-like subclasses, when presented in a well-structured way.

The objective of the third chapter is to introduce the role of rhyme and music for language processing and acquisition. First, the close link between music and language processing will be considered. As is shown by a large body of research, both domains show similar traces of neural processing. Thus, a shared neural system can, at least partially, be assumed. Second, the supportive role of music for language acquisition will be highlighted. This section focuses on the association of musical skills and language abilities. Moreover, research

underlining the advantage of the combination of music and speech for morphosyntactic learning will be presented. Finally, the relationship between rhyming abilities and language acquisition will be addressed. This part introduces research which reveals a direct link between rhyme processing and the acquisition of different language skills, such as speech production or written language abilities.

Chapter 4 focuses on a detailed description of language play and its impact on language processing and language learning. Specifically, the beneficial role of language play for the promotion of language will be pointed out. To this end, I will first provide a brief definition of language play and highlight the effect of its central features on language processing and language learning. The second part of this chapter introduces the possibility of using language play for systematic language promotion. The advantages of using language play in order to promote implicit language learning processes will be considered first. Afterwards, I will scrutinize a German didactic concept, the “Generative Textproduktion”, which uses language play for the systematic promotion of language skills.

In chapter 5, I will outline the empirical investigation of the present thesis. Based on the background provided by the preceding chapters, I will first present the fundamental questions motivating the empirical work. Second, I will give a brief overview of the three experiments, including the illustration of the artificial gender-like subclasses and their case-like paradigms. Finally, the methods, results and discussions of Experiments 1 to 3 will be detailed. Experiments 1 and 2 were conducted with German native speakers, whereas English native speakers were tested in Experiment 3. In addition to the evaluation of the separate experiments, a comparison between structured (Experiment 1) and unstructured (Experiment 2) gender-like category induction will be carried out for the German participants. Furthermore, the results of the German-speaking participants will be evaluated against those gained by the English-speaking participants.

The general discussion of the present study is presented in chapter 6. I will first summarize the main results of Experiments 1 to 3. Furthermore, the results will be discussed and related to the theoretical background of the thesis. Next, limitations of the study will be pointed out. The last section of this chapter focuses on the implications for promoting gender category induction in gender-

marking languages: How can the present findings be linked to everyday language promotion situations? This section underlines the relevance of input presentation when systematic promotion of fundamental grammar skills takes place. Finally, in chapter 7, I will provide a concluding discussion of the present thesis with regard to the main findings, the contribution to further research in this field and the implications for grammar promotion in first and second language acquisition.

2 Cues in gender-like category induction

2.1 The role of noun cues

One of the challenges of first and second language learning is to acquire morphosyntactic knowledge about the inflection of verbs and nouns. Acquiring such morphosyntactic knowledge requires the learner to find out about how individual inflectional paradigms work and which words are associated with which inflectional paradigm. Gender-marking languages, such as French, German, or Russian, typically feature different inflectional paradigms for each grammatical gender category. Table 1 presents the definite articles of German that are associated with each gender subclass (masculine, feminine or neuter singular) and case (nominative, genitive, dative, or accusative), demonstrating that each gender subclass is associated with a distinct case marking paradigm of definite determiners. Hence, a speaker must know which gender subclass any given noun is associated with so as to select the correct definite determiner. The same holds for all other morphological paradigms associated with nominal inflection, such as the indefinite articles and the inflection of prenominal adjectives.

Table 1

The Definite Articles in the German Gender-Case Paradigm (Singular Only) for Nominative, Genitive, Dative and Accusative Cases

Case	Gender Category		
	Masculine	Feminine	Neuter
Nominative	der	die	das
Genitive	des	der	des
Dative	dem	der	dem
Accusative	den	die	das

In the following, the fundamental question of how these highly complex systems are acquired will be discussed. What cues are available in speech input so that learners master the acquisition of gender categories, such as the ones

illustrated in Table 1? What is the likelihood that these different cues will occur in the acquisition of German as one's first or second language? To explore these questions, I will provide an overview of studies that have tested the impact of different cues on the acquisition of gender(-like) categories. First, I will present studies that support the effect of noun cues. Thereafter, the role of syntactic cues will be elucidated.

In a seminal study, Brooks et al. (1993) hypothesized that the acquisition of artificial gender-like subclasses might be impossible unless the nouns are phonologically or semantically marked for gender. In order to demonstrate the specific role of phonological markers for the acquisition of gender-like subclasses, they trained undergraduate students in an artificial language in two conditions. In the experimental condition, the nouns were assigned to two gender-like subclasses (Subclass 1 and Subclass 2) in such a way that their phonological form was informative regarding their assignment to gender-like subclasses. In the control condition, by contrast, the nouns were assigned randomly to Subclass 1 and Subclass 2. In both conditions, there were fifteen nouns per subclass. In the experimental condition, nine nouns of each subclass shared parts of their phonological form by ending in “-oik” (e.g. “hoik”, Subclass 1) or “-oo” (e.g. “getoo”, Subclass 2), respectively. These were the marked nouns. The remaining six nouns per subclass did not overlap phonologically and were hence unmarked (e.g. “bobell”, “billit” or “garth”). In the control group, eight nouns per subclass were phonologically marked; critically, however, both endings (“-oik” and “-oo”) were assigned to both Subclass 1 and Subclass 2 with four nouns per subclass ending in “-oik” and four ending in “-oo”. As a consequence, the markers did not provide any hint as to the assignment of a noun to one of the gender-like subclasses. Similar to the experimental group, the remaining seven nouns per subclass were unmarked.

Every noun referred to an object, such as “hoik” for house, and was combined with other elements of the artificial language to form sentences. These sentences described how an actor, “Frippy”, moved relative to the object depicted by the noun. All sentences included the agent “Frippy”, an object noun, and a locative suffix (either “eef“, “rog“, “ast“, “foo“, “ilg“, or “tev“). The suffixes indicated one of three different prepositions (to, at or from) relating to the direction or location of “Frippy” relative to the object. For instance, the sentence for “Frippy” walking towards a house (“hoik”) would be “Frippy hoik-eef”.

The suffix paradigm for the prepositions to, at, and from differed depending on the gender-like subclass of the object noun. The suffixes “eef”, “rog” and “ast” were only combined with the nouns of Subclass 1, whereas the suffixes “foo”, “ilg” and “tev” were assigned to the nouns of Subclass 2. For example, the preposition *to* translated to “eef” for nouns of Subclass 1, but to “foo” for nouns of Subclass 2. Hence, in principle, the suffix paradigm presented reliable information about the subclass of the noun with which each suffix was associated. The conditions differed, however, with respect to the availability of additional phonological cues: The majority of nouns were phonologically marked for their subclass in the experimental group, whereas systematic phonological marking was missing in the control group.

The authors worked with 30 vocabulary nouns, each of which was associated with three actions (to, at and from). Participants were exposed to the language in four or five training sessions. In the first session, they acquired the 30 vocabulary nouns. The next two to three sessions involved comprehension and production tasks regarding the artificial language. In these tasks, every sentence was combined with its corresponding action. 24 of the 90 sentences were withheld from the training and were only used in the final session. In the final session, participants were tested on two different tasks. One test was comprised of all 90 sentences for object nouns that participants had already worked with during the training sessions. These included the 24 sentences participants had not been exposed to during the training sessions. The other test worked with new material (i.e. novel nouns), examining the participants’ generalization skills. In both tests, participants were asked to produce the correct sentence associated with a given action.

The experimental group yielded significantly better results in comparison to the control group regarding both tests. This suggests that a combination of noun cues and syntactic cues is sufficient to allow participants to acquire grammatical subclasses in an artificial language. Critically, the performance attained in the control group did not exceed chance, suggesting that syntactic cues alone are insufficient for acquiring information about gender-like subclasses. In line with this, only participants who had been exposed to noun cues (seven out of 16 participants of the experimental group) were able to provide explicit knowledge about the suffix paradigm used in the study. Overall, the results suggest that phonological marking enables gender-like category

induction regarding familiar and unfamiliar nouns and supports speakers in establishing explicit knowledge about the paradigm.

In the second experiment, the authors replicated their findings with 9- and 10-year-old children. Similar to the adult study, children who learned the experimental language attained significantly better results than those trained in the control group. This held true for both familiar and novel words. However, contrary to the results seen in the adult study, only one out of seven children in the experimental group noticed that the words followed certain rules and was able to describe some explicit knowledge about the suffix system.

In a different study, Frigo and MacDonald (1998) have examined the role of phonological markers when acquiring artificial gender-like subclasses in a different domain than the locative conditions used by Brooks et al. (1993). In their study, participants learned to greet individuals in an artificial language. Similar to the experiments conducted by Brooks et al. (1993), there were two different subclasses. The subclasses referred to the status of an individual, as being either high (class 1) or low (class 2). For all individuals, participants acquired two different greeting forms (daytime greeting vs. evening greeting). Thus, in contrast to the three locative conditions used by Brooks et al. (1993), this language contained only two case-like categories. Phrases of the artificial language consisted of an individual's name and the corresponding greeting. Additionally, information about the time of day (daytime or evening) was provided. Critically, daytime and evening greeting differed with respect to the individual's subclass. That is, individuals of class 1 were greeted with "jai" (daytime) and "quo" (evening), whereas the greetings "fow" (daytime) and "mih" (evening) were assigned to individuals of class 2.

Participants were trained in one of three different language type conditions: In the systematically marked condition, 60% of the the names of the individuals of one subclass shared a phonological ending ("-ash" for class 1 and "-gor" for class 2). In the unsystematically marked condition, 60% of the names were phonologically marked, however, the marked names were evenly divided to class 1 and 2, i.e. the markers did not provide any hint as to the class membership of the name. The systematically marked and unsystematically marked conditions are comparable to the experimental and control groups in the study reported by Brooks et al. (1993), respectively. In addition to these two groups, there was a third language type condition, in which all words were

phonologically unmarked (unmarked control condition), meaning that each word had a different ending.

Participants were tested in one session, lasting about 50 to 60 minutes. They first learned the four different greetings. Then, phrase training was introduced. During the phrase training, participants conducted comprehension and production tasks. Subsequently, in the final test, participants carried out a production test with both trained and novel items.

The results indicate a clear advantage for the participants trained in the systematically marked condition. This finding applied to both trained (familiar) and novel (unfamiliar) material, but was most evident in the tests with novel items. While participants of the unsystematically marked and unmarked control conditions showed learning of some studied items, only participants of the systematically marked condition were able to generalize subclass membership to novel words. Unlike in the study reported by Brooks et al. (1993), all three groups exceeded chance regarding the trained material. However, overall, the results are consistent with those shown by Brooks et al. (1993), clearly demonstrating that phonological markers can facilitate gender-like category induction when used in different domains.

In a different set of studies, Kempe and colleagues have investigated the impact of phonological cues in a natural language, Russian (e.g. Brooks, Kempe, & Sionov, 2006; Kempe & Brooks, 2008; Taraban & Kempe, 1999). For instance, Kempe and Brooks (2008) worked with a miniature language, which was an extract of the Russian gender-case paradigm. The paradigm consisted of two Russian gender categories and three Russian case categories, similar to the paradigm applied by Brooks et al. (1993). However, due to the paradigm being an extract of the Russian gender-case paradigm, natural domains of reference of gender and case, i.e. thematic roles, were investigated here. In two experiments, the authors examined the impact of phonological noun cues in the nominative case (cues present in Experiment 1, cues absent in Experiment 2) on the learning of case-marking inflections.

Experiment 1 worked with the nominative, genitive, and dative cases of feminine and masculine Russian nouns. In the nominative case, feminine nouns ended in -a, whereas masculine nouns ended in consonants (-ø). This distinction provided phonological hints as to the gender category of the noun. Participants did not have any prior experiences with Russian or any other Slavic language.

They completed six training sessions and a test phase, which was conducted subsequent to the last training session. During the training sessions, participants first saw line drawings that either depicted an object alone (for the nominative case, e.g. a glass), or an elephant walking towards (for the dative case) or away (for the genitive case) from the object. Simultaneously, participants were aurally exposed to Russian dialogues that involved a question (e.g. “Chto eto?” (‘What is this?’)) and the corresponding answer (e.g. “Eto stakan.” (‘This is glass.’)) Next, participants worked on different comprehension and production tasks regarding the pictures and their corresponding sentences. Critically, in the production tasks in both training and test, the nominative sentence was presented first along with the associated picture. This sentence provided a phonological hint with respect to the gender category of the noun. Next, participants were asked to produce either the genitive or dative phrase for this noun. Again, as in all previous studies, participants were tested on trained and novel material.

Experiment 2 was similar to Experiment 1 with respect to the materials and procedure. As in Experiment 1, participants worked with Russian feminine and masculine nouns that were presented in three different cases. However, unlike in Experiment 1, the nominative case did not provide phonological hints as to the gender category of the nouns. That is, all feminine and masculine nouns ended in consonants. Thus, the gender of the noun was, in principle, apparent by the distributional pattern of the genitive and dative morphemes; yet, in contrast to Experiment 1, the additional phonological cue in the nominative case was missing.

The results demonstrated that learning was more successful in Experiment 1 than in Experiment 2. In Experiment 1, participants’ levels of performance were significantly better from the very beginning and increased faster than participants’ levels of performance in Experiment 2. This finding indicates that phonological cues can facilitate learning, even when these cues are transparent for one case category only. However, although participants of Experiment 2 performed poorly in gender marking, some learning of case marking was demonstrated, suggesting that participants noticed, at least to some extent, the distributional pattern of suffixes and thus the underlying case paradigm.

Taken together, the findings are consistent with the results seen in the studies reported above, suggesting a clear benefit from phonological cues in both

artificial and natural language learning. Furthermore, Brooks, Kempe and Donachie (2011) have shown that the facilitation by phonological cues reported by Kempe and Brooks (2008) can be strengthened when a larger overlap of the word endings exists, underlining the central role of phonological cues in the acquisition of grammatical categories, such as gender.

Several studies with children have demonstrated that the sensitivity to phonological cues regarding gender assignment in natural languages plays a role in processes of first language acquisition (e.g. Dabrowska & Szczerbinski, 2006; Seigneuric, Zagar, Meunier, & Spinelli, 2007). For instance, Seigneuric et al. (2007) have examined the influence of phonological and semantic cues in French-speaking children. In their study, they exposed 3- to 9-year-old French-speaking children to pseudowords that either had an ending typical for French masculine (-on, -eau, -ier, -in) or feminine (-ine, -ette, -otte, -elle) nouns, respectively, or an unbiased ending (-ige, -ale, iste, ique, -ole, -ache). The children were instructed to assign a possible gender to a given pseudoword. Here, two different conditions were administered. In the determiner condition, children were instructed to assign the gender by producing the definite article (un or une) that would most likely go with the noun. In the picture condition, children decided whether the artificial noun fits best with a female or male imaginary person by pointing to the corresponding picture.

The results showed that children gave more ending-consistent responses for the phonologically marked nouns than for the unbiased nouns, indicating their sensitivity to phonological cues when French is acquired as the first language. This sensitivity increased substantially with age. In the determiner condition, the usage of phonological cues was already noticeable in 3-year-olds, whereas the ability to allocate the correct picture to a phonologically marked noun was observable from 4 years old on. These findings suggest that, for a natural language featuring phonological cues, these cues are not only used to assign the correct gender category, but also affect semantic categorization.

Other authors used a different, connectionist approach in order to examine the validity of noun cues. For instance, MacWhinney and colleagues (MacWhinney, Leinbach, Taraban, & McDonald, 1989) implemented three computational models of the acquisition of German gender, case and plural. While the first two models examined a set of morphophonological and semantic noun cues put forward by Köpcke and Zubin (e.g. Köpcke & Zubin, 1984; see

below for an overview of the phonological cues for monosyllabic nouns), the third model did not rely on these rules. Rather, it represented the complete phonological form of each noun. In all three models, German nouns were presented to the network. Note that the nouns were taken from a corpus based on adult-directed speech and thus do not represent the input typically received by young children. The input layer represented the cues by Köpcke and Zubin (models 1 and 2) or the phonological representation of the noun (model 3). Hence, a certain activation pattern based on the presence/absence of cues (models 1 and 2) or the raw phonological form (model 3) was represented. Activation on the input layer was forwarded to the internal layers, which processed information regarding gender, case and plural. These internal layers produced activation on the output layer in the way that a German definite article was selected.

The results show that such a connectionist network can successfully use noun cues and thus learn and generalize the correct gender assignment. Model 3, using the phonological form of the German nouns, outperformed models 1 and 2 in training and generalization, demonstrating the advantage of the raw phonological form compared to the rules postulated by Köpcke and Zubin (1984). Taken together, the results suggest the central role of (phonological) noun cues in German gender acquisition. However, in order to be able to acquire the German gender assignment purely based on noun cues, it is central that the input provides these cues and the associated gender at a high frequency. As will be described in more detail below, for certain groups of children growing up in Germany (e.g. children with German as a second language or children from socially disadvantaged families), this kind of input is often not provided.

To summarize, the findings provide evidence for noun-cue models of gender acquisition, according to which phonological or semantic noun cues are indispensable for acquiring a noun's gender. Noun cues have been shown to play an important role in studies using artificial as well as natural languages in both adults and children.

In light of the findings reported by the studies above, German, a gender-marking language, presents an interesting case. In contrast to other languages, in German, a considerable number of nouns possess no phonological or semantic cues as to their gender. Although phonological and semantic cues have been identified (e.g. Köpcke, 1982; Wegener, 1995; Zubin & Köpcke, 1984),

there is a large variety of rules, and exceptions occur frequently (Wegener, 1995). For instance, Köpcke (1982) identified 24 phonological gender assignment rules for monosyllabic German nouns. However, Wegener (1995) examined the rules postulated by Köpcke (1982) and rejected them for a number of reasons. First, Köpcke (1982) worked with 1466 German monosyllabic nouns taken from a German spelling dictionary. Due to the large amount of rules allocated to a relatively small amount of nouns, the number of nouns assigned to each rule is rather small. For example, the rules assigning nouns with the initial clusters a) /tʁ/ and /dʁ/ and b) /kn/ to the masculine gender holds for 47 and 14 nouns, respectively. Second, as previously mentioned, the rules contain plenty of exceptions. Third, Wegener (1995) criticized that Köpcke did not consider the factors currentness, item frequency, and token frequency in the learners' vocabulary, since his examination was based on data gained from a spelling dictionary. Finally, 11 of the 24 rules assign two possible genders rather than one. These rules cannot be considered as distinct cues for the acquisition of the right gender, since the probability to select the incorrect gender would still be 33%. Of the remaining 13 rules that assign only one gender, 11 assign the masculine gender. Given that two thirds of the nouns Köpcke (1982) had analyzed were masculine, Wegener (1995) put forward a simpler rule that assigns the masculine gender by default to monosyllabic German nouns.

In an experiment, Wegener (1995) demonstrated that the masculine-as-default strategy can be as effective as the gender assignment rules postulated by Köpcke (1982) when allocating artificial nouns to the masculine German gender. In a different experiment, she showed that suffix rules seem to play a role in German gender assignment. The results of this experiment indicate the relevance of suffixes, such as -heit, -chen, -ung, -e or -er. Based on her data, Wegener (1995) assumed four formal gender assignment rules that play a role in first language acquisition of German (see Table 2). Beyond the masculine-as-default rule for monosyllabic nouns and nouns without a marked ending (-ø) (cf. Table 2 for examples of marked endings), she postulated certain suffix rules, with the following being most relevant for first language acquisition processes: The suffixes -el, -en, and -er assign the masculine gender whereas the feminine gender is allocated to nouns ending in -e, -ung, and -heit/keit (see Wegener, 1995, pp. 91 and Table 2).

Table 2

Formal Gender Assignment Rules Postulated by Wegener (1995)

	Gender Category
monosyllabic nouns, -ø	masculine
Suffixes	
-el, -en, -er	masculine
-e, -ung, -heit, -keit	feminine

I assessed for 222 object names taken from the Snodgrass and Vanderwart (1980) data base the predictive value of the gender assignment rules put forward by Wegener (1995). The object names had been rated previously for their age-of-acquisition (Schröder, Kauschke & De Bleser, 2003). The unambiguous assignment to the correct gender, based on Wegner's (1995) rules, was possible for 71% of all stimuli. It is noteworthy, however, that when considering the early acquired nouns that had an age-of-acquisition of 3 years or younger, this proportion dropped to 56%.

Szagun, Stumper, Sondag, and Franik (2007) have demonstrated that children do rely on some, though not all, of the available phonological cues when acquiring the German gender categories. In a longitudinal study, they investigated spontaneous speech data of 21 children with German as their first language. The authors examined the data with respect to the phonological regularities of the nouns produced by the children (e.g. endings in -er, -el, -e) and their influence on error rates concerning gender marking. The results revealed an impact of phonological regularities on the error rates when associating masculine gender. Many of the children's errors resulted from overgeneralizations of these rules. However, the results failed to show an effect of morpho-phonological noun cues when feminine gender is required. Overall, the findings indicate that children can use phonological patterns to detect probabilistic regularities, such as the ending of a word associated with a certain gender-marked article. However, this only applies when sufficient input is given and can be processed (cf. Szagun, 2004). This finding might indicate a special role for phonological cues in acquiring gender marking in German.

Indeed, Szagun (2004) showed that German-speaking children with cochlear implants with MLUs ranging from 2 to 5 were less advanced in acquiring the inflectional paradigms of the definite and indefinite articles in German than a

comparison group of normal hearing children matched for MLU. Hearing-impaired children used wrong determiners at a higher rate, with gender errors in the nominative case and omissions constituting the main error types. In contrast, the predominant error within the group of normal hearing children was the selection of the wrong case, such as the wrong usage of the accusative case “den” instead of the dative case “dem” for German masculine singular nouns. Thus, hearing-impaired children showed greater difficulties in the German gender-case system, especially regarding the correct gender assignment. Similar acquisition problems can be observed in children who acquire German as a second language (Rösch, 2003), pointing out that the acquisition of the German gender-case system constitutes a great challenge for this group of children.

To sum up, the validity of noun cues for the gender of nouns in German is limited. First, as noted previously, a large amount of early acquired monomorphemic nouns cannot be reliably assigned to one gender when phonological rules are applied. For those rules that have been postulated, exceptions occur frequently. Another question arises when considering the fact that greater than one-fifth of teenagers in Germany have a migration background (Röhner, 2005). In everyday life situations, the input for children with a migration background differs from what children with German as their native language are exposed to. It can be assumed that, at least during the first years of their language acquisition, those children receive a less frequent and less rich German input in comparison to the input children with German as their native language receive. Thus, provided that phonological cues exist, one can suppose that the extraction of these rules should be extremely hard for children who do not receive sufficient input. Research has shown that German-speaking children are able to master the gender system by the age of 3 (Szagun et al., 2007). Obviously, this acquisition process does not hold for other groups of children, such as hearing-impaired children (Szagun, 2004) or children who acquire German as a second language (Rösch, 2003). Taken together, these data suggest that phonological regularities do not represent particularly reliable cues for gender assignment in German, at least for a large amount of early acquired and frequently used nouns (Bordag, Opitz, & Pechmann, 2006; Köpcke, 1982) and especially for children who exhibit difficulties when acquiring German grammar (Szagun, 2004; Szagun et al., 2007).

2.2 The role of syntactic cues

The research presented so far has demonstrated that the gender of a German noun cannot always be reliably detected by its phonological form. Instead, the gender of German nouns appears to be more reliably indicated by syntactic cues in the morphosyntactic environment of the noun, such as the definite and indefinite determiners and other gender-marked morphemes (e.g. pronouns) (MacWhinney, Leinbach, Taraban, & McDonald, 1989; Schwichtenberg & Schiller, 2004). This is consistent with syntactic-context models, according to which the morphosyntactic context plays a pivotal role when acquiring the correct gender (Maratsos & Chalkley, 1980; Taraban, 2004). This mechanism is supposed to work especially for languages without sufficient phonological or semantic noun cues. According to the syntactic-context models, it should be possible to build lexical subclasses on the basis of correlated inflectional morphemes alone. Thus, gender categories are supposed to be learnable through the combination of the noun with small sets of associated morphemes (e.g. the determiners in the German masculine singular form (*der*, *des*, *dem*, *den*; see Table 1) plus the associated masculine singular pronouns *er* ('he'), *sein* ('his'), *ihm* ('him'), *ihn* ('him'). Apparently, children with German as their native language are successful in extracting the relevant information out of the syntactic context to correctly assign nouns to gender subclasses.

Indeed, Taraban (2004, Experiments 3 to 5) showed in an artificial language study that the acquisition of phonologically unmarked gender-subclasses is possible, if the learners' attention is drawn systematically towards the relevant syntactic context. In his study, he employed the same language as Brooks et al. (1993) but made sure that all nouns were phonologically unmarked for their gender, i.e., there were no noun cues. Unlike Brooks and colleagues, Taraban (2004) did not present full sentences, including the actor "Frippy", but used locative phrases only, such as "billit eef", meaning "ball to" in English.

In Experiment 3 of his study, Taraban (2004) examined the influence of input structuring and lexicon size on the learnability of the artificial gender-case-like system used by Brooks et al. (1993). Therefore, in one, blocked, condition, Taraban presented the input in such a way that all phrases pertaining to one noun were shown in immediate succession. For example, the three to-, at- and from-phrases for the noun "billit" (ball) were presented consecutively, allowing

participants to process the whole paradigm of locative markers pertaining to this noun. This way, their attention was guided towards the relevant gender- and case-related markers in the syntactic context. In a second, random, condition, Taraban exposed participants to all phrases in random order. This random presentation mode corresponded to the presentation mode employed by Brooks et al. (1993). In addition to this manipulation of the presentation method, Taraban investigated possible influences of the number of to-be-acquired nouns on the acquisition of gender-like subclasses: Participants were trained with either a smaller lexicon, consisting of eight nouns, or with a larger lexicon, containing 22 nouns.

Participants took part in three to five sessions, with training sessions lasting two to four sessions, carried out on consecutive days. The test session was conducted the day after the last training session. In the training sessions, participants guided their own learning by working individually on the computer until the learning criterion (94% correct translations) was reached. Unlike Brooks et al. (1993), who had presented all sentences aurally to the participants, Taraban presented all phrases visually. More specifically, an English phrase (e.g. “to ball”) was given and the participants were asked to type in the translation (“billit eef”). After each translation, immediate feedback about the correctness as well as the correct translation was given. Participants worked with the phrases in blocks of trials. After the completion of each block, the number of correct items was reported. When participants reached the learning criterion, they were instructed to return the next day for completion of the study. In the test session, the participants’ performance was tested on both training material and new material. Similar to the study conducted by Brooks et al. (1993), training material testing included phrases participants had already worked with during the training sessions (base-learning items) as well as phrases that had been withheld from the training sessions but still referred to known object nouns (base-generalization items). New material was tested by presenting one phrase of the new object noun as a hint towards its gender-like subclass (novel+hint generalization items). For example, participants were first exposed to the hint “to hat = hitab eef”, and were subsequently asked to translate the phrase “from hat”.

The results showed that the combination of a structured input with a larger lexicon facilitated learning significantly. When training occurred with the larger lexicon, participants of the blocked group scored significantly greater than chance

for the base-generalization and the novel+hint items, suggesting that they were able to generalize and use the acquired morphosyntactic knowledge. When training occurred with a smaller lexicon, participants' accuracy did not exceed chance, neither for the blocked nor for the random group. This finding suggests that a certain amount of items per category is necessary for extracting the underlying rules. The performance in the blocked group was significantly better than that attained in the random group, but only when participants were trained with the larger lexicon, indicating that blocking facilitates learning when the input contains a larger variety of lexical items.

In Experiment 4, Taraban (2004) demonstrated that when blocking was enhanced during training, blocking effects were found even when training occurred with the smaller lexicon. Here, participants of the blocked condition were only allowed to proceed to phrases for another noun when they correctly translated the set of phrases for the current noun. Additionally, the learning criterion was 100% correct responses. The results were similar to those reported in Experiment 3, with participants of the blocked condition significantly outperforming those of the random condition for both base-generalization and novel+hint items. In both Experiments 3 and 4, participants' accuracy was higher for unfamiliar material (novel+hint items) than for familiar material (base-generalization items). In order to examine this difference, Taraban (2004) carried out Experiment 5. The method was identical to the blocked condition of Experiment 4, except that the base-generalization items were also provided with a hint, similar to the novel+hint items. In Experiments 3 and 4, this hint had only been given during the novel+hint generalization test. The results demonstrate that the difference in performance between base-generalization and novel+hint items, as seen in Experiment 3 and 4, disappeared when both item types were provided with a hint.

To summarize, a blocked presentation seems to be sufficient for gender-like category induction – even if the nouns do not possess any phonological marking. Participants of the blocked condition were able to generalize and use their gained morphosyntactic knowledge and scored significantly higher than participants of the random condition. As apparent by the better results attained in the generalization tests (base-generalization, novel+hint generalization) in comparison to the training material tests (base-learning), participants of the blocked condition were able to assign the correct locative suffix when the gender-

like category was transparent through the previous hint. Taraban (2004) reports that in a related study, he used variants of the described experiments and observed that high accuracy is associated with explicit knowledge about the gender-case-like paradigm. Those participants who performed well during the test phase were able to verbalize and describe the acquired rules. To conclude, by highlighting the importance of structuring the morphosyntactic input, Taraban's findings (2004) corroborate the assumptions of syntactic-context models.

3 The role of music and rhyme in language processing and language acquisition

In chapter 1, I have considered the potential value of syntactic cues for the acquisition of gender categories. These cues might play a central role in languages that lack phonological noun cues, such as German. As shown by Taraban (2004), syntactic cues can be exploited when presented in a well-structured, systematic input. In everyday life, this organized input can be found in language play (e.g. songs or nursery rhymes). Other main features of language play are music and rhyme. Thus, the next chapter focuses on the role of music and rhyme for language processing and language learning.

3.1 Processing of music and language in adults

The questions in which ways music and language are connected and whether musical and linguistic abilities influence one another has been addressed by many studies. There are several important parallels between music and language, such as being organized temporally and involving variations for example in pitch, volume, and tone (Fonseca Mora, 2000; McMullen & Saffran, 2004). Both music and language are processed auditorily with fundamental frequency, spectral characteristics, intensity and duration as their acoustic parameters (Moreno, 2009). It is conceivable that music and language share common neural networks or substrates (Moreno, 2009). Indeed, there is some evidence that the two domains share common learning mechanisms. For instance, categorical perception, which is known to play a central role in language processing, has been found in music processing, too. Smith, Kemler Nelson, Grohskopf, and Appleton (1994) examined the perception of musical intervals and have showed that these intervals are perceived categorically. They reported evidence for categorical perception of musical material even in participants who were non-musicians[1]. Another learning mechanism shared between music and language is the sensitivity to prosodic cues, such as rhythm,

[1] Smith et al. (1994) define non-musicians as musically inexperienced novices who have not received extensive musical or instrumental training prior to the study. It should be noted that any future uses of the term “non-musicians” will refer to this definition.

stress or intonation (McMullen & Saffran, 2004). These cues serve as prosodic markers in both domains (Hirsh-Pasek et al., 1987; Jusczyk & Krumhansl, 1993).

Other studies have addressed the lexical-phonological aspects of language processing and their relationship to music processing. Using the fMRI method, Schön et al. (2010) evaluated whether lexical-phonological aspects of language and melodic aspects of music draw upon overlapping brain areas and processing ressources or whether they are processed independently. In one experiment, the authors showed that similar brain areas are activated when non-musicians process spoken words, sung words, and short melodies, with both hemispheres being involved to different degrees. In a second experiment, non-musicians listened to pairs of sung words. They were instructed to focus their attention on either the words (language task) or melodies (music task) and to judge whether the pair was identical or different regarding the relevant dimension. That is, in the language task, speech was the relevant dimension and melody was the irrelevant dimension and vice versa in the music task. Four conditions were tested: same word, same melody (1); same word, different melody (2); different word, same melody (3) and different word, different melody (4).

The fMRI analysis revealed a significant interaction between the linguistic and the musical dimension within the brain regions identified by the first experiment. Thus, the activation pattern in the relevant dimension was influenced by whether the irrelevant dimension was “same” or “different”. Post hoc tests showed that the interaction was mainly caused by the effect of the irrelevant dimension when the relevant dimension was “same”. This finding suggests that musical and linguistic information was not processed independently; instead, the irrelevant dimension influenced the relevant dimension. This applied to both words and melodies, with information of both domains influencing processing, even when they were not consciously focused upon. Overall, the results indicate that language and music influence one another and that lexical-phonological processing of language and melodic processing of music are processed interactively rather than independently.

In another study, Gordon, Schön, Magne, Astésano, and Besson (2010) expanded this design while working with a similar task. Here, non-musicians worked with the same task as the one employed by Schön et al. (2010), judging pairs of sung words regarding linguistic or musical information. While doing so,

their event-related brain potentials (ERPs) were measured. Additionally, behavioral data, namely reaction times and error rates, were collected. Again, participants judged sung word pairs while focusing either on the words (language task) or the melodies (music task).

The behavioral data showed slower reaction times and more errors when sung words were different than when sung words were the same. In the music task, the results revealed slower reaction times for different than for same melodies only when words were same but not when words were different. Also, in the music task, participants made more errors when stimuli were different than when stimuli were same for both linguistic and musical material. However, these effects were not seen in the language task. Crucially, the ERP data revealed that an N400 component, an EEG component associated with semantic violations, was elicited by both different words and different melodies. Within the N400 latency band, i.e. between 300 and 500 ms after the onset of the linguistic or musical violation, the results showed larger N400 components when targets differed from primes. This N400 effect was shown for both the linguistic and the musical material. Furthermore, the results revealed a symmetric interaction between word and melody. That is, the N400 melody effect was significant only when words were same but not when words were different. Likewise, the N400 word effect was significant when melodies were same but not when melodies were different. The results indicate that linguistic and musical information in sung words is processed interactively, which is in line with the results apparent by the fMRI method (Schön et al., 2010, see above). Several related studies confirm this neural and behavioral link between lexical-phonological aspects of language and music processing (e.g. Brown & Martinez, 2007; Brown, Martinez, Hodges, Fox, & Parsons, 2004; Kolinsky, Lidji, Peretz, Besson, & Morais, 2009).

Beyond this procedural link between language and music, there also appears to be some overlap in the mechanisms of sequence learning in both domains. Saffran and colleagues examined the sensitivity to transitional probabilities in both musical and linguistic sequence learning (Aslin, Saffran, & Newport, 1998; Mc Mullen Jonaitis & Saffran, 2009; Saffran, Aslin, & Newport, 1996a; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, & Aslin, 1996b). Such transitional probabilities are particularly informative regarding structural boundaries in both language and music.

For instance, Saffran et al. (1996b) have shown that statistical regularities help listeners to acquire implicitly the probabilities with which syllables co-occur adjacently and to use this knowledge to detect word boundaries. In their study, participants were exposed to a continuous stream of trisyllabic artificial words. Transitional probabilities, i.e. the statistical probability of the adjacent occurrence of two syllables, were higher within words than across word boundaries. Participants listened to the stream for 21 minutes in total and were instructed to find out where the nonsense words began and ended. No further information about the length, structure or amount of words was given. In the test phase, participants listened to word pairs which contained one word extracted from the language participants had been trained with. The other word was taken from an unstructured artificial language. Participants were instructed to indicate which word sounded more like a word from the training language. The results revealed that participants performed significantly better than chance with better results for those words that contained highest transitional probabilities in comparison to those words with lower transitional probabilities. Hence, participants were able to use the statistical pattern of syllables to detect word boundaries, and this ability was maximized when the pattern was most obvious (Saffran et al., 1996b).

In subsequent studies, Saffran and colleagues addressed the question whether transitional probabilities also play a role in musical structure learning (e.g. Mc Mullen Jonaitis & Saffran, 2009; Saffran et al., 1999). Saffran et al. (1999) investigated the role of statistical learning when processing and segmenting musical material instead of linguistic stimuli. Non-musicians were exposed to a continuous stream of tone sequences. Critically, the statistical structure of those tone sequences was identical to the one of the linguistic material used in the study described above (cf. Saffran et al., 1996b). That is, transitional probabilities were higher between tones belonging to one sequence than between tones belonging to two sequences. Similar to the previous study, tone sequences were presented without pauses between single sequences, thus, transitional probabilities between tones were the only cues as to the structural boundaries of the sequences. Training and test phase were similar to the procedure conducted by Saffran et al. (1996b); however, unlike Saffran et al. (1996b), Saffran et al. (1999) asked participants to avoid consciously analyzing the musical stream.

Comparable to the results seen in the study with linguistic material, participants were able to detect familiar tone sequences. Again, better results were attained for sequences containing higher transitional probabilities in comparison to sequences with lower transitional probabilties. This finding suggests that grouping and segmenting of tones is guided by transitional probabilities, even when attention is not explicitly drawn towards the task (Saffran et al. 1999). In related studies, the authors were able to show the relevance of transitional probabilities for learning of both linguistic and musical sequences in 8-month-old infants (Saffran et al., 1996a; Saffran et al., 1999). In a more recent study, Mc Mullen Jonaitis and Saffran (2009) showed that the statistical pattern of chords plays an important role for musical structure learning, even when the acquired musical structures belong to an unfamiliar music style. Saffran and colleagues postulate that a similar learning mechanism is involved when segmenting patterned stimuli in both linguistic and musical domains, indicating an overlap in structural processing between language and music.

Other studies examined processing of linguistic and musical sequences and structures and are therefore directly linked to the studies by Saffran and colleagues (see above). They provide further evidence for the overlap between structural linguistic processing and structural music processing (see, e.g., Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009 (pp. 1-3) for a recent overview). In an EEG study, Patel, Gibson, Ratner, Besson, & Holcomb (1998) found that a musical structure violation elicits a P600 component in the EEG signal, i.e. a substantial positive peak in the EEG signal 600 ms after the onset of the structure violation. This P600 had previously been demonstrated to arise after a grammatical violation within a sentence (Ousterhout & Holcomb, 1992, 1993), indicating that structural violations in music and language elicit parallel electrophysiological responses. More recently, Fedorenko, Patel, Casasanto, Winawer, and Gibson (2009) used sung stimuli to examine the relationship between syntactic processing in music and language. Participants listened to sung sentences and were asked to answer comprehension questions that followed each sentence. As for the material, linguistic and musical material was manipulated regarding their difficulty. That is, the sentences and melodies were either easy or difficult to process. Participants' accuracy for the difficult sentences was reduced when they were presented along with a difficult melody. Thus, the processing of complex linguistic structures and the processing of musical

structures interact with one another, supporting the idea of shared system for structural processing in language and music.

Moreover, an EEG study conducted by Steinbeis and Koelsch (2008) extended the studies reported previously by examining the relationship between music processing and semantic and syntactic language processing. In their study, they worked with tension-resolution patterns, which describe the structural properties of Western music, for instance regarding harmonic expectations in a sequence of chords. In a chord sequence, some chords are more likely to occur in the final position than others and are therefore perceived as more stable. This knowledge about distributional properties of music has been assumed to evolve implicitly through the exposure of Western music in everyday life (Tillmann, Bharucha, & Bigand, 2000). In their study, Steinbeis and Koelsch (2008) tested non-musicians in an interaction paradigm for investigating the impact of semantic or syntactic language violations on the processing of harmonic expectancy violations. Participants listened to sentences that were either correct or contained a violation with respect to syntax or semantics while two event-related potentials were measured (ERAN and N500). These event-related potentials have previously been shown to reflect musical-syntactic processing (ERAN - early right anterior negativity, e.g. Koelsch, 2005; Maess, Koelsch, Gunter, & Friederici, 2001) and harmonic integration (N500 - negativity maximal arousal around 500 ms, e.g. Koelsch, 2005). Sentences were presented along with a chord sequence, whereby the last chord was either a highly expected or a highly unexpected one. The main finding of the study was an interaction of harmonic structure processing on the one hand with semantic as well as syntactic language material on the other hand. This was apparent by a reduction of the event-related potentials in response to harmonic expectancy violations when presented along with a semantic or syntactic violation. Thus, beyond the link between structural music processing and structural linguistic processing, shared neural resources might also be assumed on a semantic level.

To sum up, a close relationship between language and music has been demonstrated by numerous studies. The research presented above indicates that language and music share neural resources at different sensory and cognitive levels, clearly indicating a link between both domains. Koelsch et al. (2003) demonstrated that children process music and language in a more similar way than adults. More specifically, boys between the ages of 5 and 9 tend to process

linguistic and musical structures in the left hemisphere, whereas processing was shown to be predominantly bilateral in girls. Thus, a shared system for both domains resulting from a common origin of music and language processing may well be assumed. Therefore, the question of whether language acquisition and musical learning are intertwined is highly relevant.

3.2 Music and language acquisition

Several studies have indicated that musical knowledge or musical training influences behavior and its neural substrate (e.g. the auditory cortex) (Moreno, 2009). For instance, Sleve and Miyake (2006) demonstrated that musical ability can positively influence phonological skills in second language learning. The authors tested adult native speakers of Japanese with English as their second language in several tests. The tests examined participants' L2 proficiency regarding receptive phonology, productive phonology, syntax and lexical knowledge. Additionally, in order to investigate participants' musical ability, they carried out three receptive and one productive music tests. The results indicate that musical ability might be connected with receptive and productive phonological L2 proficiency. The authors suggest that musical ability is associated with the ability to analyze new L2 sounds and is thus linked to second language phonological proficiency.

The role of musical skills in light of phonological skills has also been examined for processes of first language acquisition. Anvari, Trainor, Woodside, and Levy (2002) have examined the relationship between musical ability and phonological awareness as well as the impact of musical ability on early reading development. One hundred 4- and 5-year-olds were tested with different tasks designed to assess musical ability, phonological awareness and reading skills. The music tests evaluated processing of rhythm, melody and chord. Tests conducted on phonological awareness focused on skills that have been shown to be relevant for reading success, such as rhyme generation skills or auditory analytic skills. The results showed that musical perception skills were significantly correlated with phonological awareness, suggesting that auditory mechanism processes are shared by both domains. Since both phonological awareness and perception of music require the learner to segment and categorize units, this result is in line with studies reported above (cf. eg. Saffran et al., 1996b; Saffran

et al., 1999; Smith et al., 1994), supporting the idea of shared resources in early music and language learning. Additionally, in regression analyses, musical perception skills have been shown to predict early reading development, suggesting the important role of basic cognitive and auditory processes that underlie music perception skills (Anvari et al., 2002).

Apart from the relationship between musical and linguistic skills, musical training has been shown to enhance language abilities. More specifically, musical training can promote the development of important language acquisition skills, such as phonological awareness, auditory discrimination, reading development or the interpretation of speech rhythm (see e.g. Gromko (2005) for the effect of musical training on phonological abilities in first language learning, see Hallam (2010) (pp. 271-272) for an overview).

Moreno et al. (2009) addressed the fundamental question of whether the influence of musical skills is due to musical practice or to predispositions for musical abilities. To explore this question, the authors conducted a longitudinal study with non-musician children over a period of nine months. In their study, 32 8-year-old children were assigned to two groups and received either painting or musical training for six months. Musical training comprised training of rhythm, melody, harmony and timbre. Prior and subsequent to the training phase, children were tested in several tasks, assessing, among others, reading and speech pitch discrimination skills. The results showed that musical but not painting training had a positive effect on reading abilities and speech pitch discrimination skills. Crucially, the difference in effectiveness increased when the reading and pitch discrimination tasks were more complex. These findings suggest that musical training over a relatively short amount of time can foster the acquisition of reading abilities and pitch discrimination skills, especially when the stimuli are more complex.

This conclusion receives further support by related studies on other levels of language learning. For instance, musical training has been shown to improve verbal memory in children (e.g. Ho, Cheung, & Chan, 2003; Marin, 2009) and adults (e.g. Chan, Ho, & Cheung, 1998; Kilgour, Jakobson, & Cuddy, 2000). Ho et al. (2003) tested children between the ages of 6 to 15. Half of them had received musical training for one to five years (experimental group) while the other half received no such training (control group). The musical training in the experimental group consisted of the participation in band or orchestra programs

and additional lessons in playing classical instruments. Children with musical training exhibited significantly better verbal memory abilities than those of the control group. Furthermore, the duration of musical training significantly correlated with verbal memory skills, indicating that a longer musical training might lead to better verbal memory abilities. Moreover, those children who continued with the musical training for one year after the initial testing showed improvement in verbal memory whereas no improvement was shown in children who quit musical training within three months after the initial testing. Taken together, these findings suggest a positive effect of musical training on language skills, even if it takes place for a relatively short amount of time. However, a longer and consistent musical training is associated with better and longer-lasting effects on verbal memory skills.

The positive impact of musical training on language and memory skills has also been reported for younger children. Marin (2009) has investigated the impact of musical training on language abilities in 4- to 5-year old German children. One group of children obtained early musical training that included training with percussion instruments. The musical training was administered once a week over the time of four to five months. Language skills were recorded using Grimm's SETK 3-5 (Grimm, Aktas, & Frevert, 2001). Marin (2009) found that musical training was linked to better language abilities regarding morphological rule formation and memory for words, both of which play an important role in syntax acquisition.

Another set of studies has examined whether music can facilitate linguistic sequence learning when melody was added to a linguistic stimulus. For instance, Thiessen and Saffran (2009) exposed 6.5 to 8.0 month-old infants to strings of digit names (e.g. 9-7-3-1-5). During training, infants listened to the strings in one of two conditions: The items were either spoken (spoken condition) or sung (sung condition). Subsequent to the training phase, infants were tested with trained and novel strings. Novel strings deviated from the training items with regard to the order of the digits. Here, both familiar and novel test items were spoken, ensuring comparable conditions during the test phase. In order to assess whether the infants noticed a difference between trained and novel sequences, the authors measured the infants' attention towards each test string, using the Headturn Preference Procedure. While flashing lights were turned on and off associated with the infants' looking behavior during the training phase, a special procedure

was implemented in the test phase. First, a central flashing light was lit until the infant fixated the light. Next, a side flashing light was turned on. As soon as the infant looked at this side flashing light, the spoken test stimulus was presented. The duration of the infant looking at the light, i.e. the attention duration towards the spoken sequence, was measured. Finally, when the infant looked away from the side flashing light for more than two seconds, the sequence was stopped.

Results revealed that only infants who had previously been trained with the sung condition were able to discriminate between familiar and novel items. They paid significantly more attention to novel than to familiar sequences and were thus able to reliably detect violations present in the novel strings. This increased sensitivity seen in the sung condition suggests that melody can facilitate sequence learning. Interestingly, in a second experiment, Thiessen and Saffran (2009) demonstrated that, likewise, linguistic structure can facilitate melody learning. That is, melodies are learned more easily when presented along with language than when presented alone. This pattern of results clearly indicates the interactive effect of complex input, showing that melody learning and language learning can facilitate one another.

The positive effect of music on language learning and memory has also been shown for adults (e.g. Schön et al., 2008; Wallace,1994). Schön et al. (2008) have investigated the question of whether the link between music and language can be exploited for promoting the acquisition of transitional probability information in sequence learning. In their experiments, non-musicians listened to a continuous stream of synthesized artificial trisyllabic words that included no acoustic cues regarding the word boundaries. In order to segment individual words, participants had to detect transitional probabilities, which were higher within words than across word boundaries. The material and procedure were identical to the one used by Saffran et al. (1996b), except that the duration of speech input during the training phase was reduced from 21 minutes (cf. Saffran et al., 1996b) to 7 minutes. Also, in contrast to the study by Saffran et al. (1996b), participants were instructed to carefully listen to the stream but not to analyze it. The stream of words used in the training phase was either spoken (Experiment 1) or sung (Experiment 2). In the test phase, participants of both experiments were exposed to pairs of spoken words, each of which contained one word from the language and one unstructured, unfamiliar artificial word (nonsense-word). Participants were asked to indicate which word was most likely taken from the

language presented in the training phase. The results of Experiment 1 showed that participants were not able to discriminate words from nonsense-words when training had taken place in a spoken mode. This result contradicts the one seen in Saffran et al. (1996b), suggesting that, when reduced to 7 minutes, a spoken training input is not sufficient to detect transitional probabilities and to segment speech. However, signs of language learning were observable when participants had been exposed to a sung version of the training language (Experiment 2), with every syllable being combined with a distinct note. When melody was added to the training language in this way, participants were able to discriminate words from nonsense-words (Schön et al., 2008).

What is the mechanism underlying this finding? The facilitation of language learning through additional musical input can come about in different ways. One possibility is that the melody helps to focus attention on the linguistic stimulus and to make the input more interesting to the participant. Alternatively, the presence of tonal pitch changes may enhance the detectability of phonological boundaries and thus help to segment and discriminate sequences in the input. In order to examine whether associating each syllable with a distinct note, including syllables at linguistic boundaries, was key to the enhanced acquisition seen in Experiment 2, Schön et al. (2008) carried out a third experiment. It was identical to Experiment 2, except that the linguistic and musical boundaries no longer matched. That is, the superposition of linguistic and musical transitional probabilities was removed (variable syllable-pitch mapping condition). Participants of Experiment 3 achieved results between those seen in the spoken condition (Experiment 1) and the syllable-pitch-matching condition (Experiment 2). Critically, participants of Experiment 3 performed significantly worse in comparison to those of Experiment 2, indicating that the superposition of linguistic and musical cues was key to the enhanced language learning. However, in contrast to the results of Experiment 1, participants of Experiment 3 performed significantly better than chance and attained significantly better results than participants trained with the spoken language (Experiment 1). This latter finding supports the assumption that, beyond the positive effect of musical information as such, attention shifting and/or boundary enhancement might also play a role in language learning. Best results were attained by those participants who had received fully matched tones and syllables during training

(Experiment 2), i.e. linguistic input that associated each syllable with only one note.

Overall, Schön and colleagues (2008) conclude that music can indeed facilitate early language learning processes, such as learning to segment words. This conclusion receives additional support from a companion study, which showed that the simultaneous presentation of associated linguistic and musical structures facilitates segmentation of information in both the linguistic and the musical domain (Francois & Schön, 2010).

Taken together, these findings indicate that music might play an important role in language acquisition. The studies reviewed in this section point out three major findings. First, musical *ability* can enhance language acquisition positively. For instance, the positive transfer from musical to phonological skills and early reading development has been shown for children and adults. Second, musical *training* has been shown to improve various aspects of language acquisition, such as reading skills, phonological abilities or verbal memory. Finally, research has demonstrated that a *melodic input presentation* can directly promote language learning processes. Participants showed increased learning when linguistic stimuli were presented along with a melody compared to a non-melodic presentation. This positive effect applied for instance to the detection of phonological boundaries and word segmentation, processes that play an important role in morphosyntactic learning. Thus, music can facilitate linguistic learning through providing additional cues and directing the learner's attention towards these cues.

Jentschke, Koelsch, Sallat, and Friederici (2008) provide a different viewpoint on the relation of music and language. In an EEG study, the authors compared children with typical language development and children with specific language impairment (SLI) with respect to structural music processing. Children with SLI typically show major problems in morphosyntactic production and comprehension skills (van der Lely, 2005). Jentschke et al. (2008) exposed 4- and 5-year-old children to musical sequences while measuring EEG data regarding two-event related brain potentials (ERAN and N500). Hearing abilities and processing of acoustic features did not differ between both groups. Children listened to two types of musical sequences, each consisting of five chords. Crucially, the last chord of one sequence type was harmonically regular whereas it was slightly irregular in the other type. Both ERAN and N500 had previously

been shown to be associated with structural music processing in adults (e.g. Koelsch, 2005, see above) and children (Koelsch et al., 2003). That is, the event-related potentials are typically elicited by a structural musical violation, such as the irregular ending of a musical sequence. Both types of sequences were presented in all 12 major keys, resulting in 24 different sequences, each of which was presented eight times in total. Children of both groups (SLI children and typically developing children) listened to the musical sequences for about 17 minutes. While listening to the stimuli, children sat in front of a monitor and watched a silent movie of an aquarium.

The electrophysiological data revealed an ERAN as well as an N500 as a response to the irregular chords for the group of children with typical language development only. Thus, by detecting the structural violation, these children showed sensitivity to structural musical processing comparable to that seen in adults. In contrast, neither ERAN nor N500 was elicited in the SLI group, indicating that these children processed both types of musical stimuli in a similar way. Furthermore, correlations of the ERAN amplitude with this groups' performance in a language development test (SETK 3-5, Grimm et al., 2001) were observed. Hence, the morphosyntactic linguistic difficulties seen in children with SLI are associated with structural music processing deficits. These results are consistent with the findings of the studies reported above, clearly suggesting a relation of structural linguistic processing and structural music processing. Moreover, they might have further implications for theories of language acquisition processes and their supporting mechanisms. Based on the reported relationship of structural processing in language and music, the training of structural music processing might promote structural linguistic processing and could therefore be effective in SLI therapy. This assumption is in line with the idea that language learning may be supported by the simple and repetitive musical structure given in childrens' songs (Schön et al., 2008).

3.3 Rhyming abilities and language acquisition

The findings mentioned above indicate that musical processing, musical ability and musical training may be directly linked to language processing and linguistic skills. This has been clearly demonstrated for various language skills, such as reading abilities or phonological skills. Phonological awareness comprises

different components, such as awareness for rhymes, syllables and phonemes. Therefore, acquiring phonological abilities includes being able to detect and produce rhyme patterns. Prior research has shown that rhyme awareness and phoneme awareness can be considered as separable components of phonological awareness (Foy & Mann, 2001, 2003, 2006; Hoien, Lundberg, Stanovich, & Bjaalid, 1995; Mann & Foy, 2003).

Mann and Foy (2007) have examined the relationship between articulation skills on the one hand and awareness for rhymes and phonemes as components of phonological awareness on the other. They tested 102 4- to 6-year-old children on various tasks, including tests of rhyme awareness, phoneme awareness and articulation skills. As for the articulation skills, both intact speech production and articulatory errors, i.e. phonological processes that are atypical during language acquisition, were examined. In order to evaluate the childrens' rhyme awareness, a recognition and production task was carried out. In the rhyme recognition task, children were shown three objects, with two of them having rhyming names. The experimenter named all three objects and asked the child to pick the two that rhymed. In the rhyme production task, children were given five words and were asked to produce a word that rhymed with each given item. Here, words and nonwords were counted as correct as long as they rhymed with the item.

Children were tested on two days, with each session lasting approximately 30 minutes. After testing was complete, children were assigned to three different groups with regard to their articulation skills. Children of the Delayed Group made articulatory errors that were atypical for their age. In contrast, the Typical Group consisted of children who still made articulatory errors, but those of which are typical during speech development in their age group. Finally, the Advanced Group included children who had no articulatory deficits.

Results demonstrated a significant effect of group on phonological awareness skills; however, this applied to rhyme awareness but not to phoneme awareness. More precisely, children of the Advanced Group significantly outperformed children of the Typical Group concerning rhyme but not phoneme awareness. Also, children of the Delayed Group attained significantly worse results than children of the Typical Group for rhyme but not phoneme awareness. Furthermore, articulatory errors that were atypical in language acquisition were associated with weaker rhyme awareness whereas no relationship to phoneme

awareness was found. This pattern of results suggests a direct link between rhyme awareness and articulatory skills in preschool children (Mann & Foy, 2007).

Other findings are consistent with this result, indicating a relationship between rhyme awareness and different language abilities (Foy & Mann, 2001, 2003, 2006; Mann & Foy, 2003). Beyond the previously mentioned link between rhyme awareness and articulatory skills, the authors identified a strong association between rhyme awareness and phonological perception abilities (Foy & Mann, 2001). Using several tasks, Foy and Mann (2001) investigated the connection between phonological representations and phonological awareness in 4- to 6-year-old children. Children participated in two 45-minutes sessions. All children were preschoolers who had just acquired first formal reading experiences. Testing of phonological awareness consisted of phoneme awareness and rhyme awareness (see Mann & Foy, 2007) tests. Phonological production abilities were examined with tests evaluating performance on articulation, picture naming speed, nonword repetition and phonological distinctness, with the latter test examining the children's pronunciation ability. Phonological perception skills were tested by means of a speech discrimination task. Here, four pictures were presented, illustrating words that differed in one phoneme only. The child was asked to select the correct picture for a given word.

The results showed that phonological production skills were associated with rhyme awareness more strongly than with phoneme awareness, confirming the results reported by Mann and Foy (2007). However, the controlling factors age, vocabulary and letter knowledge critically influenced this relationship. When these controlling factors were taken into account, the relationship between phonological production skills on the one hand and awareness of rhymes and phonemes on the other was no longer present. Speech discrimination skills predicted rhyme but not phoneme awareness significantly, even when the controlling factors age, vocabulary, and letter knowledge were ruled out. (Foy & Mann, 2001). Taken together, the findings indicate that rhyme awareness is linked to phonological abilities (Mann & Foy, 2007; Foy & Mann, 2001; see also Foy & Mann, 2003).

Furthermore, there has been growing support for a connection between rhyme awareness and reading and spelling abilities. In a longitudinal study, Wood and Terrell (1998a) evaluated the phonological awareness skills of

preschool children and their reading and spelling development over the first five school terms. They initially tested 3- and 4-year-olds on a battery of phonological awareness tests assessing letter sound knowledge and segmentation, blending, alliteration, phoneme deletion and rhyme detection abilities. All children were pre-literate at that point and were about to begin with school in the upcoming fall. In the rhyme detection task, samples of four pictures were shown, three in a row at the top and one at the bottom. After each picture was introduced with its corresponding name, the experimenter instructed each child to find the picture located in the upper row that rhymed with the picture at the bottom.

After the initial assessment, the development of early reading and spelling abilities was measured over the course of two years. More specifically, all children were tested on word recognition and spelling assessment tasks at the end of their first five school terms. After completion of the study, the relationship between early reading and spelling development on the one hand and each of the initially examined aspects of phonological awareness on the other hand was investigated. Both reading and spelling development was strongly associated with prior rhyme detection abilities. Rhyme detection ability was the only variable that influenced reading development across all school terms. This influence increased constantly from the first to the last term, yielding the greatest relationship at the end of the last term. A similar picture emerged for the development of early spelling abilities, with initial rhyme detection abilities showing the strongest and most consistent contribution across all terms. The results thus suggest that rhyme awareness in pre-literate children crucially influence their reading and spelling development (Wood & Terrell, 1998a).

A different longitudinal study conducted by Bryant, McLean, and Bradley (1990) yielded results consistent with the reported findings, demonstrating that children's initial rhyme awareness skills were related to their reading abilities two years later, even after controlling for different linguistic skills, intelligence and social background. This unique role of rhyming abilities for early literacy skills has been reported in various studies as well (e.g. Hoien et al., 1995; Stainthorp & Hughes, 1998).

Rhyme awareness has also been shown to be associated with speech rhythm abilities. Wood (2006, Experiment 2) tested English-speaking children between the ages of 5 and 7 in tasks including the evaluation of speech rhythm sensitivity and rhyme awareness. Sensitivity to speech rhythm was measured

through a mispronunciation task: Children were given a picture of a house with different objects in it, each object referring to a bysillabic word. Next, children listened to words and were instructed to point to the corresponding picture. Here, they were tested in two different conditions. In the baseline condition, the given stimuli were normally pronounced. In the mispronounced condition, the metrical stress pattern of each word was reversed. Prior research had indicated that metrical stress influences language processing to a greater extent than other features related to speech rhythm (Wood, 2006, Experiment 1). As for the investigation of rhyme awareness, children conducted a rhyme detection task, in which three words were orally presented. The children's task was to identify the two words that rhymed. Rhyme detection ability was shown to be linked to metrical stress sensitivity, even after controlling for age, suggesting a link between rhyme awareness and speech rhythm sensitivity (Wood, 2006).

Rhythm in general and speech rhythm in particular are important components of different forms of language plays, such as songs or nursery rhymes. Beyond the relationship between speech rhythm sensitivity and rhyming abilities, speech rhythm sensitivity has been presumed to play a role in language acquisition processes, e.g. with regard to the identification of word boundaries (Cutler & Mehler, 1993) and the development of phonemic representations (Wood & Terell, 1998b). Based upon the assumption of a link between speech rhythm and rhyming abilities, the role of rhythm and rhyme in promotion situations using language play might be concluded.

To my knowledge, no studies have examined whether rhyming abilities might play a role in morphosyntactic acquisition processes. However, the research results discussed so far demonstrate the association between rhyming abilities and different aspects of language acquisition, including speech production, speech perception, reading and writing development and speech rhythm. Thus, one might assume that the promotion of rhyme awareness skills may also support language acquisition processes on further linguistic levels and linguistic modalities. In summary, prior research suggests that input structuring (cf. section 2.2), music and rhyming abilities can work as supporting factors in language acquisition and promotion situations.

4 Language play as input optimization

Chapters 2 and 3 highlight on the role of a) syntactic cues presented in a well-structured input and b) music and rhyme for language learning. The combination of a well-structured input with rhyming and melodic structures naturally occurs in language play. Thus, the following chapter focuses on language play and its positive impact on language learning.

4.1 Characteristics of language play

Language play can be described as a phenomenon that occurs in everyday life situations. Haueis (2007) lists a large variety of situations in which playful language is used. For instance, playful interactions between adults and young children when carrying out certain activities, such as feeding, are often accompanied by language. When considering preschool or school children, it is striking that language accompanies various playful activities, as seen in counting-out rhymes, jumping games or clapping rhymes. Different playful applications of language can, for example, be observed in bedtime monologues, in the communication among children while fictional games are played, in playful question-response sequences, riddles, texts and songs (Haueis, 2007). Although playful language is also used by adults, as seen in poetry or song, language play seems to play a crucial role for children, since various routines and playful situations are linked to language plays. In the present chapter, I will take a closer look at the definition, forms and features of language play. Therefore, first, I will give a brief summary regarding the definition of language play, followed by a discussion of the features of language play and their impact on language acquisition processes.

Language plays, such as songs, tongue twisters, nursery rhymes, poems, counting-out rhymes or clapping rhymes, may be defined as standardized, aesthetic texts that contain traditional contents and are produced in playful situations (E. Belke & Belke, 2006). Since language plays exist in various different situations of everyday life, they entail a variety of forms and functions. However, they often focus on social interactions among children or between children and adults. Characteristic forms of language play between adults and

young children are nursery rhymes and finger plays, whereas riddles, jokes, songs, counting-out rhymes and clapping games are predominant in elementary school age (G. Belke, 2012). Typically, language plays combine linguistic form with semantic content and pragmatic use, thus bringing together different levels of linguistic representation. This can e.g. be seen in functional activities that can be associated with language play, like the assignment of children to different teams through counting-out rhymes (Cook, 2000). Cook (1997) distinguishes between the formal and the semantic level of language play. The formal level contains playing with sounds (or letter shapes) in order to play with patterns, such as rhyme, rhythm or assonance, as well as playing with grammatical structures so as to create structural patterns. On the semantic level, playing and combining of semantic units takes place (Cook, 1997). The formal features of language play will be outlined in more detail shortly.

It should be noted that, when language plays are used in everyday situations, their predominant purpose is the game itself but not the acquisition of certain linguistic units or structures. However, language plays naturally draw the attention towards different linguistic units, such as phonemes, syllables and morphemes, or semantic relations (G. Belke, 2012). This is accomplished by various features of language play, e.g. repetition, rhyme or transparent structure. These features can support different language processes, e.g. the detection of syntactic structures or the segmentation of linguistic units (E. Belke & Belke, 2006), both of which are important for first and second language acquisition (cf. section 3).

Language plays exist in all kinds of cultures and societies and can therefore be considered as a cross-cultural phenomenon (E. Belke & Belke, 2006; Cook, 2000). Cook (2000) notes that there are certain childhood rhythms and melodies which are very similar across cultures. Indeed, language plays are an important form of linguistic input children receive during their language acquisition. When language play is used in everyday situations, the input is, at least to a large extent, passively received: Children are part of the communicative situation without any need to always participate productively. These situations are particularly predominant during early stages of language acquisition and thus indicate the important role of passive reception in language learning (Cook, 2000). Yet, behavioral evidence regarding the involved linguistic and cognitive processes in language play is lacking. As pointed out by Cook (2000), the major

research interest concerning interactional situations is on conversational dialogues between children and adults or among children. In contrast, the language environment of children in terms of e.g. listening to verses, songs, jokes, stories or to older children has been neglected (Cook, 2000). Due to the attention and affection language play naturally implicates, Cook (2000) postulates that this language input should be highlighted and observed more closely.

What features do language plays share that make them relevant as linguistic input in first and second language acquisition? When considering the underlying structure of language plays, certain regularities, functions and formal patterns become apparent. These features seem to be consistent through time and across different languages (G. Belke, 2007a). In the following, I will provide an overview of the central features of language play and their role for language processing and language acquisition.

Repetition. Language plays are characterized by the repetition within the texts and their unlimited repeatability. Cook (1997, 2000) reports that especially young children enjoy repeated patterns by listening to the same stories again and again without getting bored. Additionally, children like to produce language plays, such as songs, tongue twisters or nursery rhymes by heart (Cook 1997). By repeating a text over and over again, the time for language processing increases. Also, since repetition makes a text more predictable, a relaxed atmosphere is accomplished (Cook, 2000). This might be an optimal starting point for the overall memorization and storage of the text. Many language plays and children's stories repeat grammatical structures while only making few lexical substitutions, as the following examples demonstrate:

1) This little piggy went to market
This little piggy stayed at home
This little piggy had roast beef,
…

2) "What big eyes you've got grandma!"
"All the better to see you with."
"What big ears you've got grandma!"
"All the better to hear you with."
…

(cf. Cook, 1997: 229)

The repetition of grammatical structures supports the detection and isolation of certain linguistic units. Cook (2000) argues that one function of such parallelisms across repetitions might be to draw “attention to individual words, while also illustrating their occurrence in common collocations and colligations” (p. 30). Two further features of language play, predictability and regularity, are closely related to the repeatability, since all three might help to memorize words and structures (Cook, 2000). On the one hand, a text gets highly predictable due to its repetition, its regular structures and due to certain steady and simple patterns, such as rhythmic and rhyming structures. On the other hand, high predictability and regularity ensure that a text can easily be repeated, especially when repeated solely orally. This demonstrates how different features of language play can be intertwined and may influence one another.

Structured language. Language play contains a specifically structured language (G. Belke, 2012). That is, many texts show a highly systematic and transparent structure. E. Belke and Belke (2006) note that this transparent structure is due to different features, like repeated elements or phrases and a certain rhythmical structure. Typically, a limited amount of linguistic regularities structure the text. Take, for instance, a variant of a German children’s song, adapted from Fredrik Vahle’s (1988) original:

3a) Kam **der** **Igel** zu der Katze,
DET.NOM-m
' The hedgehog walked towards the cat, '

“Bitte reich mir deine Tatze!”
' “Please give me your paw!” '

“Mit **dem** **Igel** tanz’ ich nicht,
DET.DAT-m
' “I won’t dance with the hedgehog, '

denn **den** **Igel** mag ich nicht.”
DET.ACC-m
' because I don’t like the hedgehog.“ '

3b) Kam **die** **Ente** zu der Katze,
DET.NOM-f
' The duck walked towards the cat, '

“Bitte reich mir deine Tatze!”
' “Please give me your paw!” '

"Mit **der** **Ente** tanz' ich nicht,
DET.DAT-f
' "I won't dance with the duck, '

denn **die** **Ente** mag ich nicht."
DET.ACC-f
' because I don't like the duck." '

3c) Kam **das** **Schwein** zu der Katze,
DET.NOM-n
' The pig walked towards the cat, '

"Bitte reich mir deine Tatze!"
' "Please give me your paw!" '

"Mit **dem** **Schwein** tanz' ich nicht,
DET.DAT-n
' "I won't dance with the pig, '

denn **das** **Schwein** mag ich nicht."
DET.ACC-n
' because I don't like the pig." '

3a) is a variant of the first verse of Vahle's song "Der Katzentatzentanz" (Vahle, 1988). It contains the nominative, dative and accusative case for "Igel" ('hedgehog'), a German masculine singular noun (cf. Table 1). 3b) and 3c) illustrate further possible verses that contain animals with the feminine ("Ente" ('duck')) and neuter ("Schwein" ('pig')) gender (cf. Frieg, Hilbert, E. Belke, & Belke, 2012). While all lines of one verse contain the same noun, this noun occurs in the three dominant cases of German. Thus, the German gender-case system clearly structures the text while the lexical level is kept rather simple. The blocked input structure, that is the grouped presentation of the case paradigms pertaining to each noun in verses, enables the child to easily detect the grammatical associations of the determiners for the nouns.

This example demonstrates that language plays often consist of a systematic, structured input which strikingly highlights linguistic regularities. G. Belke (2012) assumes that this structured language facilitates the reception of the texts. Other prominent regularities, like rhyme, rhythm or structural parallelism can support the receptive process (G. Belke, 2012).

Regularities. Many language plays work with the violation of certain linguistic levels while other levels are maintained (Andresen, 2007). For instance,

in the next example, semantic units are substituted while the overall morphosyntactic form is kept stable (see also 3a) to 3c)).

4) When Lucy was a baby, a baby Lucy was,
She went, ‘waaa, waaa, waaa, waaa, waaa’.
When Lucy was a toddler, a toddler Lucy was,
She went toddle, toddle, toddle, toddle, toddle.
When Lucy was a schoolgirl, a schoolgirl Lucy was,
She went ‘Miss, Miss, I can’t do this.
I’ve got my knickers in a right-hand twist.’
…

(cf. Cook, 2000: 121)

Each language play contains certain regularities, reflecting e.g. grammatical rules, phonological rules or rhythmic patterns. When language plays are used as an interactive game, certain agreements are necessary in order to enjoy the text as a group. However, these agreements do not always need to be verbalized. They apply e.g. to the observation of particular regularities and the allowed violations. Only when these rules are known and observed, one can produce and reproduce the text and join the game (G. Belke, 2007a; Cook, 1997). The central role of rules can be clarified by the song “Drei Chinesen mit dem Kontrabass” (‘Three Chinese people with the contrabass’), a well-known German children’s song. The first verse, which is also the basic version of the song, is the following:

5a) Drei Chinesen mit dem Kontrabass
' Three Chinese people with the contrabass '
saßen auf der Straße und erzählten sich was.
' sat on the street and talked. '
Da kam die Polizei, fragt ‘Was ist denn das?‘
' The police came by and asked ‘What is going on here?’ '
Drei Chinesen mit dem Kontrabass.
' Three Chinese people with the contrabass. '

The subsequent verses consist of the same text, however, all vowels are exchanged by a single one, as illustrated in the following:

5b) Dra Chanasan mat dam Kantrabass
saßan af dar Straßa and arzahltan sach was.
Da kam da Palaza, fragt ‘Was ast dann das?’
Dra Chanasan mat dam Kantrabass.

When performing this language play, children use regularities in a creative way. Critically, in order to be able to participate, the children have to detect these regularities. Rules of language play can refer to the linguistic or the social-communicative level (Andresen, 2007). Hence, in this example, the linguistic regularities affect the vocalic syllable nuclei and their exchange. Critically, this requires particular knowledge about phonemes and specific language skills, namely the distinction between vowels and consonants as well as the exchange of vowel phonemes with one another. The social-communicative rules are important when the song is carried out in a group. These rules regulate e.g. the number of repeated verses or the order in which the vowels are exchanged. G. Belke (2007a) states that the development and implicit learning of these regularities is probably best achieved through the interaction in the group because in these situations, the linguistic rules are relevant to the groups' social agreements.

Creativity. Children naturally apply the linguistic regularities in a creative way. They enjoy substituting elements or transforming structures in a playful manner (Cook, 2000). This leads to the next feature of language play, namely creativity. Bell (2012) describes language play as a specific type of creative language use. She notes that the basic idea of working with language play is the manipulation of structures in order to produce new, creative sequences based upon the basic frame of the text (see e.g. 3a) to 3c)). This pattern formation has been mentioned by different authors (e.g. G. Belke, 2012; Bell, 2012; Cook, 2000).

Rhythm. A different central feature of language play is rhythm (Cook, 2000). Rhythm represents one of the crucial features in dance, music, and language, all of which are often combined in language play. Most language plays contain a simple 4/4 rhythm, which builds a clear frame due to its regularitiy and predictability:

6)	1	2	3	4
	Cock a doodle	doo, my	dame has lost her	shoe
	1	2	3	4
	My master's lost his	fiddling stick, and	doesn't know what to	do.

7)	1	2	3	4
	Round and	round the	gar	den
	1	2	3	4
	Like a	teddy	bear	P
	1	2	3	4
	One	step	two	step
	1	2	3	4
And	tickle you	under	there.	P

(cf. Cook, 2000: 19)

As apparent by 7), the beat continues during the pauses (P) and therefore the 4/4 rhythm is maintained. The speech-rhythmic breaks of language plays co-occur with linguistic boundaries and underline these to a greater extent than everyday speech does. For instance, breaks at the end of each line of the language play often co-occur with phrasal boundaries (Cook, 2000). Furthermore, in various language plays, linguistic boundaries are emphasized by subsequent motor or sensomotor activities, such as clapping on each syllable or touching a finger for each phrase. Thus, language play combines language, rhythm, and actions (Cook, 2000).

Stress pattern. As a further feature of language play, the stress pattern is directly linked to the rhythmical component. In many language play forms, stress is emphasized by the beat. As opposed to everyday speech, which naturally emphasizes stressed syllables only, language plays typically emphasize each unit of the text. Thus, in language play, also weaker syllables are highlighted. During everyday speech, these weaker syllables can easily get lost, whereas they have to be pronounced in language play in order to adhere to the given meter and rhythm (Cook, 2000). When working with language play, children are able to perceive and process all units of a text, even the ones that are hardly noticeable in everyday speech.

Rhyme and melody. Many language plays contain rhyme, which, in turn, is related to rhythm and stress. Rhyming structures are one of the most obvious features of language plays, thus drawing the attention towards the phonological form of the linguistic input, especially for young children. Moreover, rhyme helps to segment speech units, particularly when combined with a rhythm: Rhythm facilitates syllable segmentation while the segmentation of phoneme sequences is predominantly supported by the rhyming component (Cook, 2000). Melody is a further central feature of many language plays. A lot of language plays can be

sung and are directly linked to a certain melody. As previously outlined, a melodic input presentation can promote language learning (seen section 3.2).

Considering the learning mechanisms promoted by language plays, the combination of their features offers an excellent starting point for implicit learning. As previously mentioned, language plays focus on certain linguistic units or structures, at least when considering it from a linguistic point of view. Yet, this is not the child's intention when using language plays. Rather, the child seems to enjoy e.g. the rhythm, music, regularity, repeatability or the social function of a language play (E. Belke & Belke, 2006). The authors note that this is a typical situation in which implicit learning processes take place. Children primarily aim to be part of the communicative situation when producing and performing a language play. Additionally, beyond this conscious intention, they acquire or practice language without being aware of it (E. Belke & Belke, 2006). The definition and effects of implicit learning will be considered in the following.

Implicit (statistical) learning has been defined to be a "largely or wholly unconscious process of inducing structure from input following exposure to repeated exemplars" (Kidd, 2012, p. 172; see also Perruchet & Pacton, 2006). It has been deemed an automatic, spontaneous learning process, which does not depend upon declarative memory mechanisms (e.g. Forkstam & Petersson, 2005; Perruchet & Pacton, 2006; Pothos, 2007). Rather, implicit learning crucially relies on the quality and quantity of the input. Implicit learning has been documented in learning of unapparent and abstract linguistic regularities (e.g. Gomez & Gerken, 1999; Saffran, 2001; Thompson & Newport, 2007; van den Bos & Poletiek, 2010). For instance, various studies have demonstrated the impact of implicit learning on artificial grammar learning (see Pothos, 2007 for an overview). Other research has demonstrated an association between implicit learning and syntax acquisition (Kidd, 2012). E. Belke and Belke (2006) consider implicit learning as a central learning mechanism in first language acquisition, especially at early stages. Young children are already able to use their native language and its regularities. Thus, linguistic regularities are, at least partly, normally acquired at an early age. For example, children succeed in acquiring sentence structures or in assigning the correct determiner to a noun. However, they fail to verbalize their knowledge, which indicates that implicit learning has taken place (E. Belke & Belke, 2006).

Language play is typically driven by a non-verbal, semantic purpose, like the detection of a winner or the attempt to produce the text without mistakes or stumbling. Simultaneously, as noted above, it systematically draws the attention towards linguistic units and structures while no verbalization or explanation is necessary. Therefore, implicit learning might play a great role in language learning through language play. This includes the implicit learning of phonemes and structures as well as linguistic units and their segmentation.

According to this, G. Belke (2007a) mentions examples of language play forms and their benefit for language acquisition processes. For instance, the ability to segment linguistic units, such as syllables, might be facilitated by counting-out or clapping rhymes. Also, the implicit learning of syntactic structures might be strengthened through tongue twisters, which often do not only contain consonant accumulation but also syntactic inversion. Thus, language play can be considered an optimal basis for implicit learning processes (G. Belke, 2007a).

Furthermore, the pragmatic role of language play can be considered. Ritualistic actions and interactions are often associated with language play (Cook, 2000). Here, certain language plays naturally accompany different everyday activities, like jumping, cradling, eating or going to bed (G. Belke, 2012). In Germany, feeding rituals of young children are typically accompanied by structures such as the following:

8) Einen Löffel für die Mama, einen für den Papa,
' One spoon for mama, one for daddy, '
einen für die Oma aus Oklahoma,
' one for grandma from Oklahoma, '
einen für die Tante und für den Onkel auch
' one for the aunt and for the uncle, too, '
bis dem armen Kinde beinahe platzt der Bauch.
' until the poor child's belly nearly bursts. '
(cf. G. Belke, 2012: 175)

Taken together, language play combines different features, such as the repetition of linguistic units and structures, rhyme, melody and input structuring. A rhyming, melodic and structured input presentation is typical for many language plays. Despite the primarily semantic purpose of language play, the combination of the features and the associated input presentation might support the storage of words and the expansion of the lexicon as well as the acquisition of morphological paradigms (see 3) above), and world knowledge (G. Belke, 2007a,

2012). Thus, language play provides an excellent starting point for language learning, especially implicitly.

Haueis (1985, 2007) notes that playful forms of language contain an almost complete language description, including speech act models, forms of pointing and naming, the detection of syntactic structures and the segmentation of linguistic units. When considering the role of language play in language acquisition, the question of which language plays are dominant during which age is highly relevant. That is, what characteristics of language play impact at which stage of language acquisition? Another fundamental question is how language plays can be applied for language promotion in first and second language acquisition. These central questions will be discussed in the following section.

4.2 Language play and language promotion

As mentioned previously, certain features and presentation forms of language play can facilitate language learning on various levels. Language play provides a central input in first language acquisition, since young children produce and receive playful language to a large extent (Cook, 1997). There are different features of language play that are crucial at different stages of language acquisition.

The first characteristics that infants perceive from playful language are rhythm, intonation and interactional aspects, such as eye contact, touch or turn taking (Cook, 2000). At early stages of language acquisition, playful language is observable in adult-child-communication forms (Lang, 2007) and typically consist of forms like tickling rhymes (Cook, 2000). In these situations, social interactions are directly linked to rhythmic sound streams and repeated actions. Certain features of the input presentation and the associated situations, such as rhythm, stressed speech, ritualistic situations and repetition help the baby to segment sequences and allocate meaning to the constant speech stream (Cook, 2000).

In language plays for young children sound usually plays a greater role than meaning. That is, a lot of language plays preferred by young children sound interesting and contain rhyming structures whereas the meaning is unclear (Cook, 1997), as demonstrated by the following example:

9) Diddle diddle dumpling my son John
Went to bed with his trousers on
One shoe off and the other shoe on
Diddle diddle dumpling my son John.
(cf. Cook, 1997: 228)

Non-words contained in language plays do not naturally violate the structures of the language (Cook, 2000). Take for example the famous German children's song "Es tanzt ein Bi-Ba-Butzemann" ('A Bi-Ba-Butzeman is dancing'), which contains the repetition of the vowel sequence i-a-u. The same sequence can be found in the verbal tense forms of a group of German irregular verbs (e.g. singen-sang-gesungen ('sing-sang-sung')) (G. Belke, 2012).

It is well known that imitation plays a central role in early childhood in general and in language acquisition in particular. Due to its predetermined rules, language play provides a clear model for imitation, thus promoting imitative interactions (Cook, 2000; Lang, 2007). By imitating and practicing a text, the child's productive language skills are improved (Cook, 2000). Other features of language play, namely repetition, rhyme and the simple steady beat facilitate the familiarization process and thus support the child in memorizing the text. Also, children often perform over their current language ability when producing language plays. For instance, they use linguistic structures they have not yet fully acquired or unknown words (Cook, 2000). Since the use of language play is driven by entertainment and fun, this does not hinder, but might rather strengthen language learning. Taken together, language play enables children to learn their first language in the most natural way: Through repetition, rote learning and substitution, through receiving and repeating language that lacks meaning and producing language without fully understanding its meaning (Cook, 1997).

Later, at about the age of 4, children begin to actively play with language and to produce their own variants of texts (Andresen, 2007; Lang, 2007). Still, many language plays used at that age contain repetitions and variations of sound structures while the meaning is not always clear, indicating the central role of linguistic form. As the preschool children grow older, the playing with sounds is more consciously perceived, as apparent by explicit verbalizations about "nonsense words" (Andresen, 2007). At this stage, language plays are frequently observable in child-child-interaction with the pragmatic focus shifting more and more from introspection to partner aspects. For instance, children make use of parallelism and eye contact and begin to influence their partner's acting. In doing

so, explicit and implicit meta-communication is applied. Explicit meta-communication includes e.g. instructions to explicitly draw the attention towards the language play. Implicit meta-communication is used through processes like imitation, variation, integration of certain structures or consideration of timing (Lang, 2007).

Preschoolers enjoy creating their own, fictional variations of a given text. Most commonly, they substitute lexical units while the morphosyntactic structure of the basic text is maintained. As previously mentioned (cf. section 4.1), this is exactly how creative use of language play works: Certain linguistic levels are manipulated while others remain stable. Critically, children are only able to produce such creative variants if they are able to detect and distinguish between linguistic levels, indicating the learning potential of language plays. There is no need to explicitly verbalize the maintaining or violation of regularities since they can be implicitly learned and applied (Lang, 2007).

Cook (2000) proposes that the exposure to language play allows children to get in touch with literacy even before they learn to read and write. Literacy experiences prior to starting school have been shown to positively influence later reading and writing proficiency (e.g. Griffin & Morrison, 1997; Sénéchal & LeFrevre, 2002). Learning to read and write requires the child to use certain linguistic skills, like the ability to segment words (Andresen, 2007), which in turn, can be facilitated by the presentation forms of language play, such as a rhythmic and rhyming input presentation (see above). Thus, the usage of language play in preschool and early school age might play a supportive role for the development of written language abilities. Also, written language opens new possibilities for language promotion through language play, since it can facilitate the detection of linguistic regularities and paradigms. Indeed, children between the ages of 7 and 10 focus more and more on the compliance of regularities when using language play (Lang, 2007). The compatibility of language play with written language indicates the meaningful role of language play for language promotion in school settings.

The overall preference for different types and linguistic features of language play at different ages is consistent with the research conducted by Sanches and Kirshenblatt-Gimblett (1976), demonstrating that length and complexity of language play increases with age. The authors analyzed the different types of language play 301 children produced and reproduced in their

native language over the course of four years. The children ranged in age between 5 and 14 and were recruited from Austin, Texas. The sample included the following language plays: jump rope rhymes, counting-out rhymes, ball-bouncing rhymes, taunts, songs and odd verses.

The youngest, 5- to 7-year-old children produced shorter and simpler language plays. During the early school age years, the children produced more and more different language plays, holding longer and more complex texts, with most different language plays produced with 8 years of age. On average, each child produced six different language plays at this point of time. The interest in language play decreased after the age of 11, suggesting that preschool and early school age children in particular enjoy the use of language play.

The results also revealed a shift in preference for certain linguistic levels of language play with increasing age. Younger children preferred phonologically driven language plays. For instance, the only two “gibberish” language plays, which feature unique sounds without meaning, were found in the 5- to 7-year-old children. Additionally, the preference of playing with phonological units was dominant in this group of age, as apparent by the following example produced by a 7-year-old (Sanches & Kirchenblatt-Gimblett, 1976):

10) A skunk sat on a stump.
The stump said,
“The skunk stunk.”
The skunk said,
“The stump stunk.”
(cf. Sanches & Kirchenblatt-Gimblett, 1976: 103)

With increasing age, from 8 to 11, children begin to focus on grammatical, and later on semantic and sociolinguistic aspects of language play. For instance, the following rhyme, produced by an 11-year-old, plays with semantic relations (Sanches & Kirchenblatt-Gimblett, 1976):

11) Ladies and jelly-beans, hobos and tramps,
Cross-eyed mosquitos and bow-legged ants.
I come before ye to stand behind ye
To tell ye something I know nothing about.
There’s going to be a ladies’ meeting
Friday, Easter Sunday.
Men admitted free.

Pay at the door.
Pull up a chair and sit on the floor …
(cf. Sanches & Kirchenblatt-Gimblett, 1976: 103)

As the preceding paragraphs demonstrate, language play can be proposed to be a central form of linguistic input and output in first language acquisition. Children enjoy producing language plays, with different forms of language play being predominant during certain stages of age. Furthermore, as previously suggested, different features of language play and associated forms of input presentation can facilitate language learning. Given these properties of language play, they would seem to lend themselves for language promotion. When considering language acquisition in children, especially those who are at risk for language impairment or learn a second language, a fundamental question arises: What are the advantages of using language play in language promotion situations? This question is directly linked to the previously described features and presentation forms of language play and their impact on language and memory processes.

Before addressing this question, it should be noted that in Germany, multi-national classes and groups are no longer the exception but belong to our everyday life (G. Belke, 2012). Given that many children acquire German as their second language, the inclusion of systematic language promotion into everyday life situations is highly relevant. Thus, G. Belke (2012) notes that it is essential to think about optimized ways of language promotion, which should be integrated in everyday life school or kindergarten settings. As described previously, specific German grammatical forms and structures are particular difficult (e.g. the German gender-case-system, cf. section 2) and therefore constitute enormous challenges for children who acquire German as their second language (G. Belke, 2012). Hence, the systematic promotion of German grammar should be taken into account adequately. There are various reasons to include language play in language promotion and teaching situations, which will be described in the following.

First, children enjoy playing with language. Repeatability is one major feature of language play (see above) and it is well known that children enjoy receiving and producing texts over and over again. Language play combines repetition, rote learning and recitation as learning mechanisms that are presented in a playful way and therefore entertaining and enjoyable (Cook, 2000). Also,

Cook (2000) notes that, in contrast to artificial, invented examples, language play has a natural origin. It entails authentic, varied and motivated language and can therefore be considered as interesting input that induces emotions and supports discussions (Cook, 2000).

Language play further allows various interactional situations: Children can work alone, as a pair, in a small or big group. These variations make the language form and the learning topic more interesting and thus might support the child's motivation (Cook, 2000). When children work together, their social skills are promoted (Cook, 2000). The social function of language plays, such as counting-out rhymes, has been mentioned previously (cf. section 4.1). Language play combines both usage and structure of language and allows addressing and practicing them at the same time.

Another advantage of using language play in language promotion situations refers to the possibility to benefit from implicit learning processes through structured input presentation. As mentioned previously, implicit learning is a powerful learning mechanism in first language acquisition. E. Belke and Belke (2006) note that language play draws the attention towards certain phonological, morphological, lexical, grammatical and syntactic paradigms. However, the child does not primarily focus on the linguistic form, but rather on the content and result of the language play (E. Belke & Belke, 2006). Thus, language learning takes place incidentally, making language play an optimal starting point for implicit learning processes (cf. also section 4.1).

E. Belke and Belke (2006) note that implicit learning processes can be enabled through structured input presentation. A possible way to present complex linguistic input in a highly structured fashion is the usage of language play (E. Belke & Belke, 2006). Take for example the verse adapted from Vahle's song "Der Katzentatzentanz", introduced in 3a). This verse entails the three most common German cases (nominative, dative, and accusative, respectively) associated with a masculine singular noun (der Igel ('the hedgehog')). More specifically, the determiners that mainly occur with a masculine singular noun are grouped together and therefore presented in a blocked, structured way. By systematically structuring the input, children can acquire the underlying regularities, e.g. which determiners occur in association with the German masculine gender.

In the next step, children can create their own new texts. This leads to another advantage of using language play for language promotion, namely the possibility to manipulate and change a text (E. Belke & Belke, 2006). In order to change a text, the original text first has to be reproduced correctly, which requires that the specific linguistic units and structures are produced accurately. Thereafter, when the regularities are known and used correctly, the child can produce specific variations of the text. However, as discussed previously, the child has to initially know and use the linguistic regularities (e.g. the substitution of the vowels in the song “Drei Chinesen mit dem Kontrabass”, cf. 5a) and 5b), or the substitution of the noun and its determiners in “Der Katzentatzentanz”, cf. 3a) to 3c)) in order to violate them systematically (G. Belke, 2012; E. Belke & Belke, 2006). Typically, when creating a new version, the basic structure of the text is maintained while the content is manipulated through substitution of lexical units (G. Belke, 2012; Cook, 2000). For the given example, children might write verses for different animals, while the grammatical structure (nominative - dative - accusative) is maintained. By introducing other animals with the feminine and neuter gender (cf. 3b) and 3c)), children can acquire which German determiners occur together and in association with which German gender categories. Also, the acquisition of the German case paradigms can be facilitated through this blocked input presentation. Thus, language plays, such as “Der Katzentatzentanz”, enables the acquisition of the German gender categories and their functions in a systematic, structured way.

G. Belke (2007a) assumes that language play facilitates the acquisition of linguistic regularities and structures before written language is introduced. It might further promote the implicit learning of literacy and linguistic structures that are characteristic for formal language and typically occur in written language (G. Belke, 2012). E. Belke and Belke (2006) mention that the structured, blocked presentation of e.g. grammatical regularities and paradigms is especially helpful when language promotion takes place in a pre-literate stage. This way, children are able to acquire relationships between linguistic units or structures, such as the affiliation of determiners to gender categories, in a systematic way. Linguistic mental representations can be strengthened through the systematic play with language (E. Belke & Belke, 2006). Beyond the structured input presentation, G. Belke (2007a) suggests that language play facilitates language comprehension when the formal, rhythmic and syntactic pattern is predictable. Together with the

repetition of a text, these presentation forms promote the implicit learning of syntactic and morphologic structures as well as the expansion of the lexicon (G. Belke, 2007a).

Taken together, a structured input presentation along with the possibility to systematically manipulate a text represents and optimal basis for the implicit acquisition of grammatical structures and regularities, underlining that language play should be deployed when structured language promotion takes place. Further advantages of structured input presentation and creating new variants of a given text will be described in the following.

A structured input presentation predetermines certain linguistic structures and thus reduces the complexity of the language input. E. Belke and Belke (2006) propose that this given structural frame can help in creating new variants without the need to employ one's own structures. Children who begin to learn a second language tend to be especially insecure when a task requires them to produce a whole text on their own. Here, the basic frame of a language play can constitute a model which guides the output of these children. In this respect, the advantage of using language play in learner groups, especially in heterogeneous groups, is reasonable. While some children might speak the language on a high level of proficiency, other children are rather insecure and make more mistakes. Language play offers the opportunity for all children to work with a text in a creative way. Children with lower proficiencies are able to work with the basic frame and less substitutions, of which predefined solutions can facilitate this process. On the other hand, children with higher language proficiency might prefer to work more freely when creating a new text (E. Belke & Belke, 2006).

The structural frame present in a language play decreases the probability of producing mistakes. That is, when certain structures are predetermined, the error rate in the new variant can be assumed to be lower than the error rate of a task in which a new text is independently produced. This is especially relevant for second language learners. Language play allows these children to produce more correct texts and therefore enables more accurate learning (E. Belke & Belke, 2006). Syntactic priming studies with children have shown that the repetition of one structure with various contents strengthens the mental representation of this particular correct form (e.g. Shimpi, Gámez, Huttenlocher, & Vasilyeva, 2007). Furthermore, the production of a linguistically correct text is intented to boost the learner's motivation (E. Belke & Belke, 2006).

When new versions with varying contents are created, certain grammatical structures can be repeated without being monotonous or boring (G. Belke, 2012). The given frame of the language play promotes creativity and individual expression (Cook, 2000). Also, the new variants can be associated with activities, for example with a performance. Cook (2000) suggests that these activities may avoid passivity in the classroom.

G. Belke (2012) has observed that variations of language plays, especially of songs and nursery rhymes, develop naturally. Children enjoy making their own versions of a given text, demonstrating the attraction of manipulating a text for expressing their own conversational needs (G. Belke, 2012). Children enjoy this manipulation, which is, for instance, apparent by the popularity of inappropriate and indecent variations of original texts (E. Belke & Belke, 2006).

Finally, language play has been argued to foster the development of language awareness (G.Belke, 2012; E. Belke & Belke, 2006; Cook, 2000). For instance, Cook (2000) states that by drawing the attention towards linguistic regularities, one can explicitly talk about these regularities. This step does not have to contradict implicit learning processes. Rather, after implicitly introducing e.g. certain linguistic structures to the child, the explicit exposure to those structures might be useful. As previously mentioned, language play often contains structures and forms that are parallelized or grouped together (G. Belke, 2012). G. Belke (2012) notes that, as a result, the relationship of these structures and forms are obvious and easy to detect. Concluding, language play can connect the learning and reflection of a language with the practice of structures and forms (G. Belke, 2012).

To summarize, language play enables multimodal-integrated learning by connecting language with rhyme, music, rhythm, motor and sensor aspects (E. Belke & Belke, 2006). Due to its structured input presentation, language play naturally draws the attention towards linguistic units or structures. Children enjoy the manipulation of texts, which opens up the possibility of creating new variants while the basic structure of the original remains constant. This promotes implicit learning of the underlying structures or forms. Furthermore, language play can support social skills and language awareness. The preceding section clearly points out that language play can be used to promote language abilities and language learning. It can be applied at different levels of language proficiency

and to heterogeneous, multinational groups of first and second language learners (E. Belke & Belke, 2006; G. Belke, 2012; Cook, 1997, 2000).

How can we use language play in a language teaching or language promotion setting, such as in multinational school classes? This question deals with ways to integrate language plays in learner groups in order to systematically promote language learning. One German language promotion concept, the "Generative Textproduktion", applies language plays to oral and written situations to primarily promote grammatical structures, regularities and paradigms and will be addressed in more detail in the following.

Given that 25% of the younger generation in Germany acquires German as their second language (Autorengruppe Bildungsberichterstattung, 2012), the usage of effective language promotion is crucial. With this and the preceding section as background, the German language promotion concept "Generative Textproduktion" (referred to as "GT" in the following) will now be described. GT is a didactic concept for the promotion of German oral and written language abilities and was invented and further developed by Gerlind Belke (e.g. G. Belke, 2007b, 2008, 2012). It was developed in interdisciplinary classes along with literary and musical educationalists and is used by various teachers in everyday classroom settings (G. Belke, 2012).The concept can be applied to oral and written language promotion situations and thus can be carried out with preschool and school children (Frieg et al., 2012). It is typically used with groups of children, such as in everyday teaching situations at school or in special language promotion settings. However, it can be applied to individual language promotion and speech therapy situations as well (Frieg et al., 2012). GT is thus mainly used in language relevant teaching and promotion situations. Additionally, since textual topics can be adjusted to specific thematic areas, GT can also be integrated in subject teachings, such as mathematics or science classes (Frieg et al., 2012).

G. Belke (2012) notes that appropriate materials for working with GT are texts that are relatively limited and contain clearly marked linguistic forms, as often seen in language plays. The focus of GT is the creative production of new, personal variants of a given language play (G. Belke, 2012; Frieg et al., 2012). Language plays entail attractive texts that can be easily memorized. The repetition of language, a central requirement for successful language learning, can be ensured in the absence of monotony (G. Belke, 2012).

As the foregoing section has demonstrated, language plays seem to be particularly interesting for preschool and elementary school children (Sanches & Kirshenblatt-Gimblett, 1976). Indeed, G. Belke (2007a) notes that GT is mainly used with elementary school children. She further proposes that the concept is well-suited for the integration in everyday life school settings, as the usage of language play is attractive for children with both German as their first and second language. As mentioned previously, language play enables the specific promotion of standard language units and structures, which is required for a successful acquisition of written language. Furthermore, language play facilitates a new way of reflecting upon different relationships, such as the relationships between the norm and creativity or between systematic and spontaneous learning, learning processes that both L1 (first language) and L2 (second language) learners have to master. Finally, language play might foster the connection of literary learning and language learning and is therefore beneficial to all children (G. Belke, 2007a).

GT enables implicit learning of linguistic regularities that are central for the successful acquisition of written-language abilities. The acquisition of written language is a key task of institutional learning with the purpose of expanding orally acquired language proficiency with respect to written-language norms (G. Belke, 2012). Overall, the aim of working with GT is to implicitly acquire linguistic structures through texts that clearly highlight the relevant structures. When working with GT, G. Belke (2012) recommends focusing on few grammatical structures that are considered difficult to acquire for second language learners of German, such as the German gender assignment or the acquisition of the different verb conjugations in German.

The basic procedure when working with GT can be divided into two different stages. First, children produce and reproduce a given text. Thereafter, they create their own new variants, based upon the original text (G. Belke, 2012; Frieg et al., 2012). When creating new variants, children automatically use structures from the original text. This usage of the basic frame facilitates the production of linguistically correct texts (G. Belke, 2012; Frieg et al., 2012). Furthermore, the repeated exposure of certain structures in different variants promotes an intuition for textuality and syntactic coherence as well as implicit learning of the presented structures (G. Belke, 2012). G. Belke (2012) states that the younger the children, the easier they implicitly acquire complex linguistic

structures. The memorization of linguistic structures and words is facilitated through repetition and variation of language plays, while there is no need to explicitly verbalize the learning processes (G. Belke, 2007b).

In the following, a more precise description of a possible procedure when working with GT will be given (cf. Frieg et al., 2012; Hoffmann & Weis, 2011): First, the text is presented to the children through reading or singing it aloud. This is repeated until the children are able to produce the text, ideally until it is memorized. The memorization leads to a better understanding of the text and thus to the expansion of the lexicon. The memorization further leads to improved pronunciation skills and a more accurate and fluent production of the original text. While repeating the text, active elements, such as rhythmic-motor accompaniment or the acting out of a text, can be included. The inclusion might facilitate the comprehension of the content and the memorization of the linguistic forms (Frieg et al., 2012; Hoffmann & Weis, 2011). When working with children who can read already, the written text can be given additionally. Children can read the text aloud, for example as a group or in alternation with a partner, which links oral experiences to written forms (Frieg et al., 2012). As mentioned above, this repetition of a text is considered the first stage of the basic procedure of GT. In the following, the second stage, namely the creation of new variants, will be described.

Second, first variations of the original text can be produced as a group. When new elements are substituted, the focus of attention is automatically drawn towards the original structures and their changes. These instructed variants often inspire children to create their own versions (Frieg et al., 2012). Hoffmann and Weis (2011) note that beyond the playful and creative act of practicing the language, the repetition and variation leads to a highly structured input presentation of linguistic structures. Yet, children primarily focus on the content of the different versions rather than on the language, which, along with the optimized input presentation, is a well-suited starting point for implicit learning processes of linguistic structures (Frieg et al., 2012).

For children who can already read and write, the implicit learning processes can be combined with explicit learning, by consciously guiding the attention towards linguistic structures and promoting children to ask questions. That is, through substituting certain elements, children are able to easily detect the relationship between the substitutions and the associated linguistic changes

(G. Belke, 2012; Frieg et al., 2012). For instance, the substitution of different animals in the verses of “Der Katzentatzentanz” (cf. 3a) to 3c) above) highlights the changes of the determiners and thus the regularities of the German gender-case system. When children detect those regularities, GT can promote language awareness. This awareness of e.g. grammatical regularities can be illustrated and retained in tables and posters. Color- or bold-marking can help to identify and structure linguistic relationships and changes (G. Belke, 2012; Frieg et al., 2012).

The illustrations mentioned above can render assistance when children create new variants of the original text, which constitutes the third step of the GT procedure. In this step, children produce their own versions of the basic text. Children use the given frame of the original in order to create new variants (Frieg et al., 2012). They develop new variants by substituting linguistic elements. The substitution can be either content- or structure-based. For the content-based substitutions, certain units (e.g. words, phrases) are substituted while the basic linguistic structure is maintained. Structure-based work can be defined as a transformation of structures (e.g. active voice into passive voice) by which the core content remains unchanged (G. Belke, 2012; Frieg et al., 2012).

The initial substitutions should be conducted closely to the original text. For instance, the first degree of variation in the verse given in 3a) is to substitute the German masculine noun “Igel” (‘hedgehog’) with other German masculine nouns, such as “Löwe” (‘lion’) or “Hund” (‘dog’). This does not yet require any changes to the morphosyntactic context of the noun, since both “Igel” and “Löwe” are masculine. Thereafter, animals with different genders, i.e. feminine and neuter nouns (cf. e.g. 3b) and 3c)), can be substituted, requiring morphosyntactic adaptations (Frieg et al., 2012). A major advantage of producing new variants based upon an original text is that only few systematic changes have to be made in order to produce one’s own, personal text. New variants can be created solely orally or also in written form. The usage of written language is often associated with linguistic regularities being brought to awareness (Frieg et al., 2012). Frieg et al. (2012) propose a control phase as the fourth step, in which the accuracy of the new texts is examined.

As for the material selection, G. Belke (2012) suggests different linguistic focus areas for different stages of age: In elementary school, the acquisition of simple morphological structures through substitution of single elements should be

practiced intensively, so that children are able to correctly use declensions, conjugations and function words. Later, in secondary education, complex syntax, i.e. passive structures, and thus more and more transformational changes have to be considered (G. Belke, 2012). Practical experience shows that children handle the creative phases of GT differently depending upon their language proficiency and their thematic interest. While some children adhere closely to the predefined substitutions, others vary the text more freely or are inspired to write a completely new text (G. Belke, 2012).

Hoffmann and Weis (2011) assume that not only language plays, but also childrens' books can be used when working with GT. Childrens' books are well-suited for language promotion when they contain the prominent and/or repetitive presentation of the crucial linguistic structure, since this motivates children to produce their own parallel texts (Hoffman & Weis, 2011). Indeed, von Lehmden, Kauffeldt, Belke, and Rohlfing (2013) have demonstrated that the frequent presentation of a certain grammatical structure in childrens' books leads to better comprehension and production skills of this structure when book reading takes place in an interactive fashion.

To date, there are only few articles in the literature in which the impact of GT is considered. Arslan (2005) has examined the successful use of the concept in language promotion of second language learners. In special language promotion sessions, GT was introduced to elementary school children who acquired German as their second language. Prior to the promotion phase, all children were able to comprehend German accurately. However, their German language production abilities were characterized as weak in terms of oral and written language, including severe difficulties with the German determiners. The language promotion thus focused on the systematic practice of the German determiners and their inflectional forms. During the language promotion settings, different language input was presented, such as childrens' books, songs and games. GT was introduced through various songs, which were frequently repeated, modified, and acted out throughout the sessions. One song was performed in front of an audience. Children wrote their own verses based upon texts that contained the German determiners in a highly structured presentation, comparable to the examples given in 3a) to 3c). By producing their own verses, the children were able and motivated to write additional new verses and stories, thus further strengthening the correct usage of the German determiners. At the

end of the language promotion phase, the children wrote texts that were altered and deviated further from the original text. These texts demonstrated a better and largely accurate usage of the determiners in comparison to the texts written prior to the promotion phase (Arslan, 2005).

Similar conclusions were reported by Frieg et al. (2012). The authors found a positive effect of GT on language skills, when deployed in speech therapy settings. More precisely, two 5- and 6-year-old children with German as their second language received speech therapy sessions that included GT in oral and written form. The authors observed increased grammatical abilities and lexicon expansion at the end of the therapy phase (Frieg et al, 2012). These observations suggest that GT might effectively facilitate language abilities when used in language promotion and speech therapy settings. Further research is necessary to evaluate the effect of the concept when applied to everyday school settings and when used with other groups of children, such as children with specific language impairment.

In summary, the foregoing section demonstrates that language play can be effectively used for promoting language abilities in first and second language learners. GT, a didactic concept for multinational groups, focuses on the systematic language promotion by means of working with language play. It can be applied to various oral and written situations, such as language teaching in school, language promotion for groups of children or individual speech therapy. The concept has been shown to facilitate grammar acquisition in children with German as their second language, when used in language promotion and speech therapy settings. GT combines optimized, structured language input with the pleasure and enjoyment of using language play and hence builds an outstanding basis for implicit language learning.

5 Empirical investigation

5.1 Motivation and questions of the present study

In the preceding chapters, language processing and language learning have been reviewed and discussed from different viewpoints. First, prior research has examined what cues can help facilitate the acquisition of complex grammatical forms and structures. As discussed above, the German gender acquisition constitutes a major hurdle for many learners of German, such as second language learners. Taraban (2004) has demonstrated that syntactic cues, if structured systematically, are sufficient for gender-like category induction when noun cues are lacking. The optimization of syntactic cues, i.e. a blocked input presentation, can facilitate the acquisition of gender-like subclasses. Given that many monomorphemic German nouns possess only few semantic or phonological noun cues about their gender, the role of syntactic cues for the German gender acquisition seems to be of central relevance.

Second, former research has addressed the relationship between music and language. Beyond a link between music and language processing on different levels, different studies have shown the positive impact of music on language abilities (e.g. Moreno et al., 2009) and during language learning (e.g. Schön et al., 2008). Third, the potential value of rhyming abilities (e.g. Wood, 2006) for language acquisition processes has been explored.

Taken together, prior research has demonstrated that input structuring, music and rhyme can influence language acquisition positively. Language play typically combines a structured input presentation with rhyming and/or melodic structures. More precisely, language play often contains certain linguistic structures in a highly structured and structurally repetitive way. That is, syntactic regularities as well as grammatical changes are clearly marked. This distinguishes the input of language play from everyday interaction, which is structurally rather variable. Additionally, rhyming and melodic structures are often given in language play and might facilitate language memory and language learning.

The advantage of using language play for language promotion has been pointed out above (e.g. Cook, 2000; G. Belke, 2012). In light of the fact that about

25% of the children in Germany acquire German as their second language (Autorengruppe Bildungsberichterstattung, 2012), the question of how effective language promotion and language teaching can be conducted, especially for preschool and elementary school children, is highly relevant. Several authors have noted that children with German as their second language show particular problems with the acquisition of the German gender categories and the associated case paradgims (e.g. Jeuk, 2007; Kaltenbacher & Klages, 2006; Rösch, 2003). For instance, Jeuk (2007) reported that the correct use of German determiners constitutes a major problem in elementary school children who learn German as a second language, with children overgeneralizing, omitting or unsystematically using certain forms.

G. Belke (2007a) proposes a didactic concept which considers teaching and learning of German in light of first, second and foreign language acquisition. She recommends common teaching of German for all children which takes into account the language difficulties of all children. Jeuk (2007) also suggests that German teaching in elementary school should be adapted with respect to a stronger focus on language promotion. As for the need of systematic and structured language teaching, grammar teaching should comprise the work with linguistic forms that are clearly marked in written language (Jeuk, 2007). As previously mentioned, those forms are especially obvious in language plays (cf. section 4.1). It has been argued that the combination of the presentation forms of language play (input structuring, rhyme, melody) render language plays and excellent starting point for language learning (E. Belke & Belke, 2006; Cook, 1997, 2000; Haueis, 1985). GT, a didactic concept for multinational learner groups, focuses on language promotion through language play. Thus, it combines the previously described positive effects of the features of language play (input structuring, rhyme, melody) with language promotion that naturally occurs in preschool and school settings. The present paper intends to verify this claim of using language play for language promotion empirically.

The purpose of the present study is to examine how the combination of the presentation forms input structuring, rhyme, and melody affects the acquisition of gender-like subclasses in an artificial grammar that provides exclusively syntactic cues towards the gender-like subclass of any given noun. Based on language play as one possibility to promote grammar skills, my research questions were the following:

1. Can the acquisition of gender-like categories be facilitated when optimizing the input structure?

With this question I aimed to examine whether I am able to replicate the results reported by Taraban (2004). Thus, the focus of my interest was on finding out how the exposure to linguistic input can be optimized in such a way that the assignment of nouns that are phonologically unmarked for gender can be acquired more easily. Similar to Taraban's (2004, Experiments 3-5) experiments, in the blocked training method, sentences were presented in a systematically structured, blocked fashion, thus optimizing the syntactic cues. In contrast, in the random training method, sentences were presented in a random, unstructured fashion. My objective was to determine whether such facilitation by input structuring is still effective when the artificial language features three (as in German, see Table 1) as opposed to two (as in Brooks et al., 1993 and Taraban, 2004) subclasses. Given that many children in Germany acquire German as their second language, the focus of the study was on investigating *implicit* ways of facilitating the acquisition of artificial gender-like subclasses, which can, in principle, be applied to young children.

2. Do rhyme and melody, which are both important components of language play, affect the acquisition of the gender-like subclasses?

With this question I aimed to investigate to what extent adding rhyme, melody or both to the training material impacts the acquisition of gender-like subclasses in the artificial language. Linking both questions 1 and 2, this was assessed with two modes of input structuring, a highly structured presentation (Experiment 1 and parts of Experiment 3 in the following) and a random, unstructured presentation (Experiment 2 and parts of Experiment 3 in the following).

If the forms of input presentation that naturally occur in language play support the acquisition of grammar skills, structured input learning should show greater learning improvement in comparison to random learning. Additionally, differences between learning with and without the combination of rhyme and melody might be expected.

5.2 Overview of the experiments

In three experiments, the role of the presentation forms input structuring (through blocking), rhyme, and melody when learning artificial gender-like categories was examined. The artificial language contained three gender-like categories and three case-like, locative categories (see Table 3). Unlike any of the preceding studies, there was considerable overlap in the gender-case-like markers featured within and across the inflectional paradigms associated with each gender. Take for example the marker *ino* (see Table 3) seen in the to-locative of gender-like category I. It is also used in the by- and the from-locatives of gender-like category II. This overlap is similar to the one associated with the German definite article *der*, which occurs in the nominative case and its masculine singular form as well as in the feminine singular forms of the genitive and dative cases (see Table 1). Such syncretism is characteristic of the German gender-case system, which fills a total of twelve grammatical functions in the singular inflectional paradigms with only six different forms (see Table 1).

Table 3

The Artificial Gender-Case-Like Markers Used in Experiments 1, 2 and 3

Case-like Category (Locative Markers)	Gender-like Category		
	I	II	III
To-Locative (I)	ino	ano	ono
By-Locative (II)	gla	ino	gla
From-Locative (III)	sla	ino	sla

Since German native speakers have already acquired the German case paradigms during their first language acquisition, the artificial language should comprise a different domain than case. Thus, following Brooks et al. (1993), the German case paradigm was replaced by a paradigm of locative markers (expressing to, by, and from). There were no noun cues as to the gender-like category membership of the nouns. Therefore, when creating the artificial nouns (see Experiment 1, section 5.3.1), the item pool included bisyllabic words that did

not possess any of the most relevant gender-indicating suffixes postulated by Wegener (1995) (see Table 2).

In the first two experiments, German native speakers learned to comprehend and produce sentences in a blocked (Experiment 1) or random (Experiment 2) sentence presentation. In Experiment 1, all sentences concerning one noun were presented in immediate succession. That is, the sentences associated with one noun were grouped together. This way, participants could mutually process all three case-like markers that belonged to one noun. In Experiment 2, all sentences were presented in random order so that the sentences were no longer grouped together for each noun. In both experiments, participants were trained in one of the following four modes: a) prose, b) rhyme, c) melody, and d) rhyme and melody. In Experiment 3, English native speakers were tested with selected methods of the language. Based on the results obtained in Experiments 1 and 2 (see Experiment 1, section 5.3.2 and Experiment 2, sections 5.4.2 and 5.4.3), Experiment 3 included the training of four blocked groups (as in Experiment 1) and one random group (as in Experiment 2)

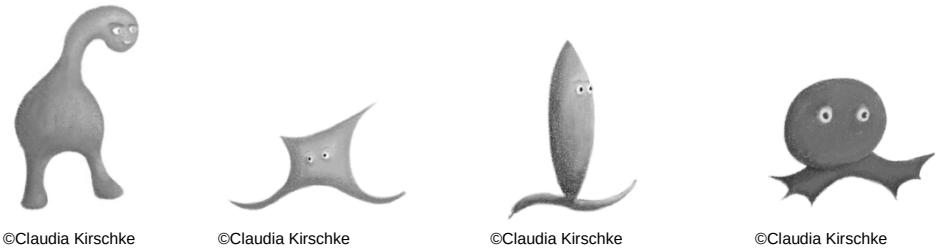
5.3 Experiment 1: Structured gender-like category induction in German native speakers

5.3.1 Method

Participants. 48 participants, recruited at Ruhr-University Bochum and University of Münster, Germany, took part in exchange of payment or course credits. All participants were German native speakers between the ages of 18 and 30 years. They were randomly assigned to one of four groups, with each group consisting of 12 participants.

Materials. For the language training, animated scenes in which phantasy creatures moved in relation to each other were created in Microsoft PowerPoint. These animated scenes were combined with sentences describing the actions. In each scene a single creature, *Tika* (see a) in Figure 1), moved either towards, around, or away from a variety of other phantasy creatures. Creatures b) to d) in Figure 1 provide examples for three of these other creatures. All pictures

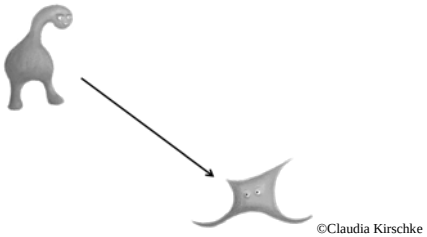
depicting the phantasy creatures were created by Claudia Kirschke, who has kindly made the pictures available to us. As will be explained in more detail shortly (see section *The language*), the animated scenes showed *Tika* walking towards, around, or away from another creature. This is exemplified in Figure 2.



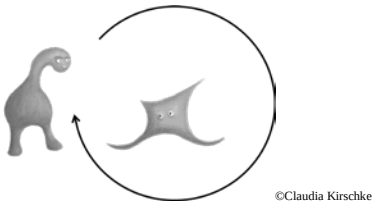
a) Tika b) Lelop c) Filan d) Tekes

Figure 1. Examples of the fantasy creatures used in Experiments 1, 2 and 3.

a) to-locative (ino)



b) by-locative (gla)



c) from-locative (sla)

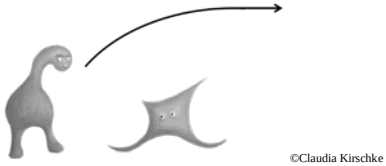


Figure 2. Examples of the animated scenes of Tika's movements towards (a) around (b) or away from (c) „Lelop". For „Lelop", the correct grammatical markers associated with each locative were *ino*, *gla*, and *sla*, respectively.

The lexicon. All elements of the lexicon were pseudowords in German and English with no orthographic German or English real-word neighbors. The pseudowords were generated using the program WinWordGen 1.0 (Duyck, Desmet, Verbeke, & Brysbaert, 2004). The pseudowords had a high summated bigram frequency in both German and English and were easy to pronounce for native speakers of both languages. Despite their status as pseudowords, I shall refer to the individual lexical entries as nouns, verbs and markers, in order to indicate their grammatical function in the artificial language.

The lexicon of the artificial language contained 37 nouns, all referring to one fantasy creature. 36 of the nouns formed the grammatical function of an object, such as „Lelop“*, “*Filan” or “Tekes” (see Figure 1). These nouns will be referred to as *object nouns* in the following (see Appendix A for an overview of all object nouns and their features). Every sentence contained the one subject noun, *Tika*. In addition, two verbs (*pim* and *pif*) were included in the lexicon. These verbs ensured that all sentences differed from one another. *Pif* referred to the action of Tika moving towards or away from another creature, *pim* referred to Tika moving around another creature in a circle. Lastly, the lexicon contained five markers that gave information about the gender- and case-like classes of the object nouns (see Table 3).

Each object noun consisted of four to five letters (see Appendix A). They were pronounced as disyllabic words with four to six phonemes. I ensured that none of the object nouns included suffixes associated with a specific gender in German (e.g., -e, -el,-en, -er, -heit and -ung, see Wegener, 1995).

Each of the 36 object nouns depicted one fantasy creature that interacted with the one subject noun creature (Tika). The creatures differed in their shape and color and were designed to not include any gender-specific features. In a norming study, 21 native speakers of German (students of Ruhr-University Bochum) were asked to inspect the creatures and to spontaneously judge whether they considered the creatures to be male, female, or neuter by selecting the definite article that they would spontaneously associate with each creature. Subsequently, I presented the same participants with audio recordings of the German pronunciations of the creature nouns and asked them to note for each stimulus how they thought it would be written in German and which definite article (der, die, das) would most likely go with it. Those creatures and object nouns that

scored higher than 50% with one specific gender (e.g. with the masculine gender) were equally assigned to all three gender-like categories of the language. This way, an equal distribution of the critical object nouns was ensured. Subsequently, the remaining object nouns were also allocated randomly to all three gender-like categories resulting in groups of 12 object nouns per gender-like category.

There were five grammatical markers carrying information about the gender- and case-like categories of the object nouns (see Table 3). As detailed previously, this set of markers differed from the sets of markers used in previous studies in that only five markers fulfilled a total of nine grammatical functions within the gender-case-like paradigm. The systematicity of the artificial system was therefore similar, at least partly, to the German gender-case system.

Since German native speakers have already acquired a gender-case system in their first language, the system of gender-case-like markers did not relate to thematic roles as the German determiners do. Instead, the artificial markers expressed three locative conditions. Here, the choice of grammatical markers was determined by the locative orientation of the subject noun's (Tika's) action and the respective gender-like subclass. That is, a sentence was associated with a to-locative marker when Tika walked towards an object noun creature, it was associated with a by-locative marker when Tika walked around an object noun creature, and it was associated with a from-locative marker when Tika walked away from an object noun creature (see Figure 2). In addition to the locative conditions, every object noun was assigned to one of three gender-like subclasses (I, II or III, see Table 3). Thus, e.g. for „Lelop", an object noun creature belonging to the gender-like category I, the markers were *ino* (to-locative), *gla* (by-locative) and *sla* (from-locative) (see Figure 2).

The language. The markers were embedded in sentences, using the following syntactic structure: First, the object noun was produced, followed by the locative marker, the subject noun, and the verb [object noun – marker – subject noun (Tika) – verb]. The sentence structure differed from the standard German and English sentence structure in the order of the syntactic units. The positions of the object noun and the marker were also reversed in order to avoid an association of the artificial language with the standard German or English syntactic structure. Due to the three case-like categories, each object noun was associated with three sentences expressing that Tika walks towards, around, or

away from the object noun creature, respectively. Thus, for example, for the creature „Lelop“, which was assigned to the gender-like category I, the three sentences are shown in (1) to (3). In the glosses, G refers to the respective gender-like category (I-III) and C to the respective case-like category (I-III).

(1) to-locative

Lelop	*ino*	*Tika*	*pif.*
Lelop.OBJ	ino.GI.CI)	Tika.SBJ	walk.towards.PRS.3SG

'Tika walks towards Lelop.'

(2) by-locative

Lelop	*gla*	*Tika*	*pim.*
Lelop.OBJ	ino.GI.CII)	Tika.SBJ	walk.around.PRS.3SG

'Tika walks around Lelop.'

(3) from-locative

Lelop	*sla*	*Tika*	*pif.*
Lelop.OBJ	ino.GI.CIII)	Tika.SBJ	walk.away.PRS.3SG

'Tika walks away from Lelop.'

For the language training, Microsoft PowerPoint presentations were prepared, showing the subject noun creature, Tika, walking towards, around, or away from an object noun creature. For the vocabulary training, an additional set of 31 digital and printed slides as well as vocabulary cards, both showing the creatures plus the orthographic representation of their associated nouns, were used. The slides and cards depicted the creatures that were used during the training phases (see section *Procedure* below).

Design. Experiment 1 employed a blocked input presentation, that is, all sentences concerning one noun were presented consecutively. Thus, for example, participants were first exposed to all three sentences associated with the creature „Lelop“ (see sentences (1) to (3) above), then to all three sentences associated with “Filan”, and so on. The sentences were presented in the following sequence: Tika walking towards the object noun creature, Tika walking around the object noun creature, and Tika walking away from the object noun creature.

Language learning took place in one of four different modes, resulting from the combination of the training variables rhyme (present, absent) and melody (present, absent). In Group B-P, the sentences were neither presented in a rhyming nor a melodic structure. This mode will be referred to as “prose”, with the B in the group’s initial referring to the blocked presentation of the training material described above. Group B-R was exposed to all sentences in a rhyming version, hearing all sentences followed by an additional nonsense rhyming line. These lines rhymed with the last syllable of the sentence and were the same for all object nouns, as exemplified below for „Lelop“:

(4) to-locative
Lelop ino Tika pif. (sentence)
Biri bi, biri bif. (rhyme)

(5) by-locative
Lelop gla Tika pim. (sentence)
Bim bam bim, bim bam bim. (rhyme)

(6) from-locative
Lelop sla Tika pif. (sentence)
Biri, biri, biri, bif. (rhyme)

Group B-M was exposed to a melodic version of the training material – here, all sentences were presented in a sung version. Finally, for Group B-R&M, the sentences were presented in a rhyming and melodic version.

Since the presentation modes of Groups B-M and B-R&M were created with reference to childrens’ songs, I made sure that the structures of the sung sentences fit with standard structures of such songs. Based on the existing sentences, a six-lined structure was given for Group B-R&M while a three-lined structure occurred in Group B-M (see examples for „Lelop“ above). Since a six- or four-lined structure (but no three-lined structure) is common in childrens’ songs, a four-lined structure for Group B-M was created by embedding a nonsense line after the second sentence (doobee-daa, doobee-doo). When creating the two melodies (for Groups B-M and B-R&M), they were matched according to their form, structure and musical quality.

The purpose of the melody was to maximally support the important grammatical information, namely the markers and their combination across

sentences. Thus, the focus was on bringing together the components melody, grammatical structure and text content in an optimal way. From a musical point of view, childrens' songs lend themselves well for achieving this goal. For instance, they feature redundant melodies with repeated notes. Prior research indicates that the use of repeated notes can enhance the memory for melody (Schellenberg & Trehub, 1999; Trainor & Trehub, 1992). Hence, the melodies composed for the present material were based upon elemental common features of popular music, like repeatability, the use of major and minor harmonies and simple rhythm. These features, which are typical for childrens' songs, support the quick comprehensibility and accessibility of popular songs (Kramarz, 2007; see also Kauffeldt, 2011).

Procedure. Participants took part in three training sessions, followed by one test session. Out of the 36 object nouns, 30 were used for training purposes. In the first session, the participants learned the names of the 31 creature nouns, namely the 30 object nouns and the one subject noun (vocabulary training). In sessions 2 and 3, they were trained with the sentences in reception and production tasks. In session 4, they were tested in several tasks. The sessions were conducted sequentially within two consecutive weeks, one session lasting approximately 60 minutes. In each week, two sessions were carried out. Participants were instructed that they would learn an artificial language and that no further information about the language could be given until the debriefing after the completion of the study. In the following, first the procedure of the training sessions and then the test procedure will be described.

Training procedure. At the beginning of *Session 1,* participants learned that they would get to know 31 fantasy creatures and were supposed to learn their associated nouns as a requirement for the next sessions. This session included the extensive training of 30 object creature nouns (nouns 1 to 30 in Appendix A) plus the subject creature noun (Tika). Participants were exposed to the creature nouns in five blocks, each block containing six to seven creatures and their corresponding nouns. Each block was trained separately. First, the items were introduced by presenting each picture of a single creature twice along with its spoken and written names. Subsequently, only images of the creatures were shown and participants were asked to name each of them from memory. Immediate auditory and orthographic feedback followed each response. This task

was continued until participants were able to name all creatures of the block of creatures correctly. When the learning criterion was met, participants proceeded to the next vocabulary block. After completing all five vocabulary blocks, participants were given a set of cards, showing all 31 creatures and their written names. They were instructed to review them at home and to return the cards to the experimenter at the beginning of the next session, which took place two or three days later. Apart from the written creature names used in Session 1, no further orthographic information was provided in the present study. All linguistic information was provided aurally so as to ensure that the training method paralleled the conditions of first language acquisition as closely as possible.

During all following training sessions, 6 of the 30 acquired object nouns were withheld from further training sessions in order to be used as novel-old items in the test session (nouns 25 to 30 in Appendix A). They were only featured in the vocabulary review carried out at the beginning of sessions 2 and 4 to ensure that participants knew the names of all 30 object creatures. Thus, apart from the vocabulary review sections, each training session was comprised of three tasks pertaining to 24 of the 30 object nouns. As each object noun was associated with three actions (see Figure 2 and examples (1) to (3) for „Lelop“ in section *Materials*), all tasks in the training sessions contained 72 sentences in total. During all tasks, participants were told that they should give the best answer they could, even if they were not sure about the accuracy of their response.

Session 2 started by testing the knowledge of all 31 creature nouns that had been acquired during the first session. All participants reached the learning criterion of 95%, i.e., 29 correct responses. The same applied to participants of Experiments 2 and 3. The second part of session 2 involved two receptive language learning tasks. The training began with a task requiring participants to listen to all sentences while watching the related actions (blocked listen-and-repeat task, Task 1). As described above, all 72 sentences were presented in a blocked fashion. For each trial, participants listened to the three sentences associated with one object noun and were then asked to repeat them. Afterwards, participants listened to the correct sentences as auditory feedback to monitor their response. Subsequently, in Task 2 (sentence-action matching), the participants listened to 24 single sentences, each referring to one object noun. Each sentence was presented along with a sample of pictures showing the three

associated actions of the object noun creature (see e.g. Figure 2 for „Lelop“). Participants were asked to select the corresponding image. The participants received immediate feedback about the correctness of the response via a red box marking the correct picture. There was no blocking by object nouns in Task 2, as there was only one test item per object noun.

After accomplishing the receptive tasks, productive language training was introduced. In Task 3 (blocked sentence production), participants were instructed to watch all 72 actions (as in Task 1) again. However, this time, they were asked to produce the corresponding sentence for each action. Again, all actions were presented grouped by object nouns; however, unlike in Task 1, participants produced the sentence for each action immediately and did not wait for the three actions for a creature to be completed. After each group of three sentences, the correct sentences were given as auditory feedback.

In *Session 3*, participants first completed the blocked listen-and-repeat and sentence production tasks once again (Tasks 1 and 3). After that, they worked through Task 4 (random-order sentence production) which was similar to Task 3 in that participants watched all 72 actions and were asked to produce the corresponding sentences. In contrast to Task 3, however, the actions were no longer grouped together but presented in random order. Auditory feedback, again in the form of the correct sentence, was given after each sentence. Task 4 was similar to Test 1 (carried out in the test session, see below), except for the auditory feedback given in Task 4, which was absent in Test 1. By including it as the very last task in the last training session, I was able to compare participants' performance on this task immediately after training and in the test session, which took place two or three days after the last training session. Thus, short-term versus long-term learning for the random-order sentence production task was examined.

Test procedure. The final session began with the vocabulary review of the 31 creature nouns. In the first two tests, all stimuli pertained to the 24 object nouns the participants had worked with during Tasks 1 to 4 (tests with old items).

Test 1 (random-order sentence production) was identical to Task 4, except that participants received no feedback and were exposed to only half of the trained material. To be precise, the material used here comprised the three actions pertaining to 12 of the 24 trained object nouns (four nouns per gender-

like subclass), resulting in 36 actions in total. As in Task 4, participants watched the actions and were asked to produce the associated sentence.

During Test 2, participants performed a forced-choice grammaticality judgment. They watched 36 randomly assigned actions, taken from the remaining 12 trained object nouns that had not been used in Test 1. All actions were combined with an aurally presented sentence. The participants were instructed to categorize the sentences regarding their grammatical correctness by assessing whether the sentence matched the action. The sentences were either correct (one third of all stimuli) or contained an incorrect marker in terms of a) the gender-like category (one third of all stimuli) or b) the case-like category of the noun (one third of all stimuli). No feedback was given. To perform successfully in Tests 1 and 2, knowledge about the gender-like class of each object noun as well as the right application of the gender-case-like rules was necessary.

In the remaining two tests, participants worked with new, untrained sentences in order to test participants' ability to generalize the acquired regularities to new object nouns. To this end, the tests employed the six object nouns that had not been included in the training tasks, but had been reviewed regularly in the vocabulary review (novel-old object nouns, see Appendix A). Additionally, six novel-new object nouns were tested that the participants had never seen or heard of before (see Appendix A). Two novel-new and two novel-old object nouns were assigned to each gender-like category.

Test 3 (novel+hint sentence production) required participants to generate sentences related to actions performed with the novel object nouns. Since participants could not infer the gender-like category of the novel nouns, first, a hint was given by aurally presenting the to-locative sentence of an object noun along with the corresponding action. Next, the participants were asked to generate the two remaining sentences (by-locative and from-locative) for the object noun along with the visual presentation of the actions. Again, no feedback was given.

Finally, in Test 4 (novel forced-choice grammaticality judgement), participants listened to sentences pertaining to novel vocabulary. The sentences were presented in groups of three that each pertained to the same novel object noun. Thus, one group of sentences contained the to-, by- and from-locatives for a novel noun, respectively. For each of the 12 novel object nouns, one correct and one incorrect group of sentences was presented, resulting in 24 groups in

total. Incorrect groups (50% of all groups) included one incorrect sentence, which was either false concerning a) the gender-like category, b) the case-like category or c) both, with each type of error featuring in one third of the incorrect sentences. Participants were instructed to decide whether the group of sentences associated with one noun sounded grammatically correct or incorrect. No feedback was provided.

After the tests, the participants filled out a questionnaire. A copy of the questionnaire is attached in the Appendix (see Appendix B). It included questions about the language background of the participants and about the level of awareness regarding the artificial gender-case-like paradigm. For instance, participants indicated their experience with foreign languages and answered questions about the structure of the acquired material to test the retrieval of the learned grammatical system.

5.3.2 Results and discussion

All tests were analyzed with respect to marker accuracy. In the production tests (Task 4, Tests 1, and 3), the proportion of correct markers produced in all valid utterances were analyzed. Responses were coded as errors if participants had used the wrong markers. Other mistakes, namely the wrong use of an object or verb, occurred rarely (15 out of 6336 cases, 0.24%) and were not excluded from the analyses since the correct marker application was still possible despite the wrong use of a verb or noun. In the grammaticality judgment tests (Tests 2 and 4), the proportion of correct responses was assessed. There were no invalid trials, that is, all participants responded to all stimuli.

Prior to the statistical analyses of the effect of the training group (prose, rhyme, melody, rhyme and melody) on the proportions of correct responses, the data was arcsine-transformed. I carried out both analyses by participants (reported as F1 for all analyses of variance), with the independent variable training group as a between-participants variable, and analyses by items (F2), treating training group as a within-items variable. For post hoc tests, I computed Tukey's HSD for the participant analyses and paired t-tests for the item analyses. Post hoc analyses will be reported as significant if the analyses showed *p*-values $< .05$.

In a first set of analyses, I compared the participants' accuracy with the chance level for Task 4 and Tests 1 to 4 using one-sample t-tests.[2] The results showed that the accuracy of all four groups significantly exceeded chance in all tests ($p < .05$ by participants and by items), indicating that all groups had learned the gender-case like system of locative markers.

To examine the influence of training group (B-P, B-R, B-M, B-R&M) on the participants' accuracy, I carried out one-factor ANOVAs for tests conducted on old items (Task 4, Test 1, Test 2). For the generalization tests (Tests 3 and 4), I conducted ANOVAs including training group and familiarity of novel object nouns (novel-old, novel-new) as independent variables.

[2] Since there were only five markers, the chance level to choose the correct one is 20%. However, due to the blocked method of training, it was expected that participants would quickly pick up on the locative marker paradigm for each gender-like subclass (ino-gla-sla, ano-ino-ino, ono-gla-sla, see Table 3). Indeed the data confirms this assumption (see the results of Test 3 reported below). The only error participants would commit in this case would be to choose the wrong paradigm, that is, to assign the wrong gender-like category to a given object noun. Therefore, the chance level was set at 33.3%.

Table 4

Experiment 1 with German Native Speakers (Blocked Presentation in Training): Percentages of Correct Markers in Task 4 (Random-Order Sentence Production with Old Object Nouns at the End of the Last Training Session), Test 1 (Random-Order Sentence Production with Old Object Nouns), Test 2 (Forced-Choice Grammaticality Judgment with Old Object Nouns), Test 3 (Novel+Hint Sentence Production, Collapsed Over Novel-Old and Novel-New Nouns) and Test 4 (Novel Forced-Choice Grammaticality Judgment, Collapsed Over Novel-Old and Novel-New Nouns)

Training Group	Old Items			Novel Items	
	Production		Comprehension	Production	Comprehension
	Task 4	Test 1	Test 2	Test 3	Test 4
Blocked Prose	71.99 (16.73)	60.88 (16.43)	77.31 (8.36)	92.36 (19.29)	96.18 (6.52)
Blocked Rhyme	80.44 (16.46)	75.23 (20.01)	82.18 (11.87)	92.36 (10.93)	97.22 (2.71)
Blocked Melody	83.22 (13.14)	82.64 (14.48)	87.27 (12.50)	99.31 (2.41)	99.31 (1.62)
Blocked Rhyme & Melody	92.48 (10.62)	90.28 (19.43)	96.06 (5.09)	100 (0)	99.65 (1.2)

Note. Standard deviations are given in parentheses.

Old items (Task 4, Test 1, Test 2).

Production tests. In order to consider short-term learning vs. long-term learning, the random-order sentence production test was conducted at the end of the last training session (Task 4), as well as in the test session (Test 1). Participants watched randomly assigned actions and were asked to produce the corresponding sentences. All sentences had been trained during the training sessions. The percentages of correct markers produced in these tests are presented in Table 4.

In Task 4, performance increased from Group B-P to Group B-R&M. There was a main effect of training group, $F1(3, 44) = 4.98$, $p = .005$; $F2(3, 213) = 48.61$, $p < .001$. Post hoc tests (by participants and items) showed that performance was significantly higher in Group B-R&M compared to Group B-P.

The by-items analyses revealed several additional significant differences between groups (B-P < B-R, B-P < B-M, B-R < B-R&M, B-M < B-R&M) that did not reach significance in the post-hoc analyses by participants. This indicates that rhyme alone and melody alone improved the performance of some, but not all, participants in groups B-R and B-M.[3] It is clear from the significant by-items analyses that the few participants that performed differently than the other participants did so with respect to all items, suggesting that they had learned the underlying gender-like category assignment.

In Test 1, too, there were significant differences between training groups, $F1(3, 44) = 6.99$, $p = .001$; $F2(3, 105) = 34.54$, $p < .001$. Post hoc tests (by participants and items) revealed significantly better results of Groups B-M and B-R&M in comparison to Group B-P. Again, the by-items analyses revealed additional significant differences (B-P < B-R, B-R < B-M, B-R < B-R&M, B-M < B-R&M). Hence, even after some time had passed since the last training had elapsed, group B-R&M still outperformed all other groups. As before, results revealed that rhyme and melody alone facilitated learning in some but not all participants, with group B-M faring overall better than group B-R.

Comprehension test. Table 4 presents, for each training group, the percentage of correct responses in the forced-choice grammaticality judgment test, performed with training material (Test 2). In Test 2, participants watched randomly assigned actions, which were each accompanied by an aurally presented sentence. The participants were asked to judge whether the aurally presented sentences matched the actions. All groups achieved higher levels of correctness in the comprehension test than in the production tests. The ANOVAs revealed a main effect of training group, $F1(3, 44) = 10.34$, $p < .001$; $F2(3, 105) = 22.41$, $p < .001$. Post hoc tests (by participants and items) showed that Group B-R&M attained significantly better results compared to all other groups. The by-items analyses established an additional significant difference between the groups B-P and B-M. Again, this data suggests an advantage of melody alone regarding some, but not all participants.

[3] It is unlikely that the non-significant results seen in the corresponding post hoc tests of the by-participants analyses are due to a lack of power. After all, the analyses are powerful enough to detect the difference between groups B-P and B-R&M. However, it is noteworthy that training group is a between-participants but within-items factor, which is why subtle effects may be detectable in the by-items but not the by-participants analyses.

To summarize the findings from the tests with old items, the results from the production tests (Task 4, Test 1) demonstrate that while rhyme alone and melody alone leads to improvement in some participants, the most consistent improvement is seen when rhyme and melody are combined (Group B-R&M). This group's advantage was visible even when the task was carried out immediately after training and performance of all groups exceeded 70%. In the comprehension test (Test 2), Group B-R&M clearly stood out again, significantly exceeding the performance of all other groups. These findings indicate that the blocked training method is efficient, even in marker systems as complex as the one used in the present study, and even if training is performed aurally instead of visually (cf. Taraban, 2004). Furthermore, the significant increase in learning performance in Group B-R&M indicates that this training mode can facilitate gender-like category induction when added to a blocked presentation.

Generalization tests (Tests 3 and 4). With Tests 3 and 4, one production and one comprehension test with novel items was carried out. Recall that participants had not worked with the sentences associated with the novel nouns. In order to perform well on these tests, participants had to generalize and use their knowledge about the subclass assignment rules.

Production test. In the novel+hint sentence production test (Test 3), participants worked with novel-old and novel-new object nouns. Novel-old object nouns referred to creatures that had been included in the vocabulary training in session 1 and the vocabulary review in sessions 2 and 4 but had not featured in any other training task. Novel-new object nouns referred to creatures that participants had never seen before (see Appendix A). For each novel creature, participants first watched a to-locative action and listened to the corresponding to-locative sentence. Next, they were asked to produce the by-locative and the from-locative sentences for this creature while watching the corresponding actions. This way, they produced two sentences per novel object noun. In the statistical analyses of the arcsine-transformed proportions of correctly produced sentences, training group and familiarity were included as independent variables, coding for each sentence whether it featured a novel-old (familiar) or a novel-new (unfamiliar) object noun.

Table 4 shows marker accuracy for Test 3 per training group, collapsed across novel-old and novel-new object nouns. The statistical analyses showed

that the degree of familiarity with the novel nouns did not impact on participants' accuracy and did not interact with teaching mode (all *F*s < 1).

As apparent by Table 4, all blocked groups attained excellent results. Nevertheless, Groups B-M and B-R&M still clearly excelled and performed better than all other groups. The main effect of group was significant in the by-items analyses and approached significance in the by-participants analyses, *F1*(3, 44) = 2.32 *p* = .088; *F2*(3, 66) = 31.20; *p* < .001. The post hoc tests by items showed that Groups B-P and B-R yielded significantly worse results than Groups B-M and B-R&M and that there were no differences between Groups B-P and B-R or between Groups B-M and B-R&M. A descriptive analysis by participants showed that 39 of the 48 participants attained 100% correct. 8 of the 9 participants who did not score 100 percent had been trained in Groups B-P or B-R, mirroring the findings from the analyses by items.

Comprehension test. Finally, participants carried out the novel forced-choice grammaticality judgment test with six novel-old and six novel-new object nouns (Test 4). To reiterate, they were asked to judge whether the three sentences associated with a novel object noun were grammatically correct or not. As mentioned previously, for each novel object noun, one correct and one incorrect group of sentences was presented. As with Test 3, training group and familiarity (novel-old, novel-new) were included as independent variables, coding for each set of sentences whether it featured a novel-old (familiar) or a novel-new (unfamiliar) object noun. Participants were more accurate when responding to sentences pertaining to unfamiliar (99%) than to familiar (97%) object nouns, *F1*(1, 44) = 5.75, *p* < .05; *F2*(1, 22) = 1.24, *p* = .278. Critically, familiarity did not interact with training group (both *F*s < 1.6, *p* > .208).

Table 4 presents the percentages of correct responses for this test, collapsed across both types of novel nouns. The participants of Groups B-P and B-R fared slightly worse than the participants of Groups B-M and B-R&M. Even though the differences between groups were rather subtle, the main effect of group was significant, *F1*(3, 44) = 2.86, *p* < .05; *F2*(3, 66) = 3.25, *p* < .05. Post hoc analyses showed significant differences between individual groups in the by-items analyses only, with Group B-P yielding significantly worse levels of performance than Groups B-M and B-R&M. Like in the novel+hint sentence production test (Test 3), the levels of performance achieved in the novel

grammaticality judgment test were extremely high, exceeding 95% correct responses in all four groups.

Taken together, all groups attained excellent levels of performance in both generalization tests. Clearly, the participants were capable of using and generalizing the learned rules, indicating that rule-based learning had taken place. Even though the participants performed near ceiling, there were still differences between the groups with outstanding results for Groups B-M and B-R&M. The advantages for these training groups were apparent by the by-items analyses only, indicating that not all, but some participants of these groups attained excellent results and that those participants did so across all items. Participants of all groups attained excellent results in both generalization tests, demonstrating that they were able to use the marker assignment rules and thus had successfully acquired the underlying paradigm. It can be assumed that rule-based learning was not only used in the generalization tests, but was also, at least partly, applied in the tests with old items (Task 4, Tests 1 and 2).

At first sight, it may seem odd that participants performed better in the tests with novel items (generalization tests) than in the tests with old items (see Table 4). However, in all tasks with old items, participants worked with individual sentences that were presented in random order and they received no hints regarding the gender-like subclass of the individual object nouns. Thus, to perform well on the tests with old items, participants needed to use not only the paradigms of locative markers but also assign the right gender-like category to each object noun. Therefore, provided that participants were aware of the rules of marker assignment associated with each gender-like category, the generalization tests (Tests 3 and 4) were easier to complete than those working with old items (Task 4, Tests 1 and 2).

In summary, the results from Experiment 1 demonstrate that a blocked presentation is sufficient for the acquisition of gender-like subclasses. Participants of all groups used the acquired regularities concerning both trained and untrained material. Moreover, participants of Group B-R&M attained the best results in all tests. The findings demonstrated a clear advantage for Group B-R&M regarding both training material (tests with old items) and new material (generalization tests). This group's performance stood out especially in the tests with old items, which required the participants to not only use the marker paradigms but also to assign the correct gender-like category to a given noun.

Overall, the results indicate that a combination of rhyme and melody in training the blocked materials enhances the learning effect considerably.

5.4 Experiment 2: Unstructured gender-like category induction in German native speakers

The purpose of Experiment 2 was to explore to what extent the artificial gender-case-like paradigm was learnable when the training sentences were presented in a random fashion instead of a blocked one. Additionally, the present experiment examined whether the different presentation modes (prose, rhyme, melody, rhyme-and-melody) affected learning in a similar way when combined with a random training method.

5.4.1 Method

Participants. There were 48 participants in this experiment. They were recruited and compensated in the same way as participants in Experiment 1.

Materials and design. The materials were identical to those used in Experiment 1. In contrast to Experiment 1, the sentences for each object noun were no longer grouped together but were presented in a random order. Hence, for example, first the to-locative sentence for “Filan” was presented, next the from-locative sentence for „Lelop“, then the from-locative sentence for “Tekes”, and so forth. As in Experiment 1, participants were trained in one of four different training groups to examine the influence of the variables rhyme and melody. These groups will be referred to as R-P (prose, no rhyme, no melody), R-R (rhyme, no melody), R-M (no rhyme, melody) and R-R&M (rhyme and melody), with the first R in each group’s name denoting the random-order presentation of the training material in these groups.

Procedure. Participants were randomly assigned to one of the four training groups. The procedure was identical to Experiment 1, except that the training sentences were presented in a random fashion. It is noteworthy that in the listen-and-repeat tasks (Task 1) administered in Experiment 2, participants were no longer required to memorize and repeat three sentences in a row, as it

had been the case in the blocked training method used in Experiment 1. Instead, they repeated each sentence immediately. If anything, this should ease the processing load associated with memorizing the individual sentences, enhancing participants' training outcome.

5.4.2 Results and discussion

The data analysis was identical to that used in Experiment 1. Again, the use of marker accuracy was analyzed. Sentences containing the wrong use of an object or verb occurred more often than in Experiment 1 but were still rather rare (265 out of 6336 cases, totaling 4.18% of the data). As in Experiment 1, they were not excluded from the analyses. When comparing the participants' accuracy to chance, the chance level was, again, set at 33.3% in order to compare the results to the one found in Experiment 1 (see Footnote 2). One-sample t-tests showed that participants' marker accuracy was significantly greater than chance ($p > .05$) for all analyzed tests, suggesting that learning had occurred in all four training groups. This result is not compatible with the finding reported by Brooks et al. (1993) who had found that, with a random presentation of the material in a prose mode, participants did not exceed chance level. However, Brooks et al. (1993) did not arcsine-transform their data. When carrying out the analyses with untransformed data, the R-P and R-R groups did not exceed chance level regarding Test 1, replicating, for the R-P group, the findings reported by Brooks et al. (1993).

Table 5

Experiment 2 with German Native Speakers (Random Presentation in Training): Percentages of Correct Markers in Task 4 (Random-Order Sentence Production with Old Object Nouns at the End of the Last Training Session), Test 1 (Random-Order Sentence Production with Old Object Nouns), Test 2 (Forced-Choice Grammaticality Judgment with Old Object Nouns), Test 3 (Novel+Hint Sentence Production, Collapsed Over Novel-Old and Novel-New Nouns) and Test 4 (Novel Forced-Choice Grammaticality Judgment, Collapsed Over Novel-Old and Novel-New Nouns)

Training Group	Old Items			Novel- Items	
	Production		Comprehension	Production	Comprehension
	Task 4	Test 1	Test 2	Test 3	Test 4
Random Prose	54.75 (17.94)	49.54 (19.56)	66.67 (9.77)	50.69 (17.48)	69.79 (10.97)
Random Rhyme	62.73 (15.27)	55.32 (20.67)	69.91 (9.31)	61.46 (19.39)	76.39 (8.94)
Random Melody	60.07 (16.92)	58.33 (16.15)	74.07 (9.79)	64.24 (19.98)	72.92 (15.64)
Random Rhyme & Melody	65.28 (19.14)	60.65 (18.83)	75.23 (10.95)	64.58 (26.62)	82.64 (13.74)

Note. Standard deviations are given in parentheses.

Old items (Task 4, Test 1, Test 2).

Production tests. As noted above, the random-order sentence production test was carried out at the end of the last training session (Task 4) and in the test session (Test 1). Results of both tests are summarized in Table 5. For both tests, the main effect of training group reached significance in the by items-analyses but was not significant in the by-participants analyses (Task 4: $F1(3, 44) = .902$, $p = .448$; $F2(3, 213) = 8.34$, $p < .001$; Test 1: $F1(3, 44) = .752$, $p = .527$; $F2(3, 105) = 3.77$, $p = .013$). For Task 4, post hoc tests by items showed that Group R-P yielded significantly worse results than the three remaining Groups R-R, R-M and R-R&M. Furthermore, Group R-M attained significantly worse results in comparison to those of Group R-R&M. For Test 1, levels of performance increased from Group R-P to Group R-R&M. Post hoc tests

by items revealed significantly worse results of Group R-P in comparison to those attained by Groups R-M and R-R&M. All of these differences between groups were caused by only a few of the participants in each group, as indexed by the non-significant results of the by-participants analyses.

Comprehension test. The results of the forced-choice grammaticality judgment test (Test 2) are presented in Table 5. As seen with the random-order sentence production tests, analyses revealed a main effect of training group in the by-items analyses only, $F1(3, 44) = 1.69$, $p = .183$; $F2(3, 105) = 4.98$, $p = .003$, with Group R-P yielding significantly worse results than Groups R-R, R-M and R-R&M.

Overall, the tests working with old items revealed that Group R-P performed worst and Group R-R&M best. This finding is line with the results seen in Experiment 1, suggesting an advantage of being trained with a combination of rhyme and melody. However, in contrast to Experiment 1, the present results were due to significant effects of the by-items analyses only, indicating that only a few participants showed the described effects.

Generalization tests (Tests 3 and 4). As detailed previously, the production and comprehension generalization tests were carried out with six novel-old object nouns, which referred to creatures that participants had worked with during the vocabulary training but in no other part of the training phase. Additionally, six novel-new object nouns were introduced, which participants had never encountered before.

Production test. Table 5 presents the test results of Test 3 (novel+hint sentence production), collapsed across novel-old and novel-new object nouns. ANOVAs including familiarity (novel-old, novel-new) and training group as independent variables yielded no effects of familiarity or its interaction with group (all Fs < 1). The main effect of group reached significance in the by-items analyses only, $F1(3, 44) = 1.51$, $p = .226$; $F2(3, 66) = 2.86$, $p < .05$. The post hoc tests by items showed that Group R-P attained significantly worse results than Groups R-M and R-R&M.

Comprehension test. Table 5 summarizes the results of the novel forced-choice grammaticality judgment (Test 4), collapsed across both novel object noun types. ANOVAs including familiarity (novel-old, novel-new) and training group as independent variables yielded a main effect of familiarity in the

by-participants analyses, $F1(1, 44) = 12.50$, $p < .01$; $F2(1, 44) = 1.11$, $p = .303$. Participants yielded higher levels of accuracy when producing sentences with unfamiliar object nouns (79%) than with familiar object nouns (72%). Critically, familiarity did not interact with group (both Fs < 1.8, p >.492).

The main effect of training group was significant in the analyses by items and approached significance in the analyses by participants, $F1(3, 44) = 2.58$, $p = .066$; $F2(3, 69) = 4.83$, $p < .01$. Post hoc tests by participants and by items showed that Group R-R&M attained significantly better results than Group R-P. Further post hoc tests by items showed that Group R-R&M yielded significantly better results than Group R-M and that Group R-R attained significantly better results than Groups R-P and R-M.

To summarize, Group R-R&M attained best and Group R-P worst results in all tests, demonstrating a similar pattern to that seen in Experiment 1. It is striking, however, that, except for the novel forced-choice grammatical judgment (Test 4), the main effect of group reached significance in the by-items analyses only. This suggests that not all participants of the R-R&M group improved during training but that those who did, did so for all items. Clearly, training participants with the random training method yielded less consistent results than training participants with the blocked training method, which had induced significant effects of group in both the analyses by participants and the analyses by items (see Experiment 1). Hence, it would appear that training mode (prose, rhyme, melody, rhyme-and-melody) had a weaker but otherwise quite similar effect on the participants trained with the random training method (Experiment 2) compared with the participants trained with the blocked method (Experiment 1): The combination of rhyme and melody leads to best results and might thus facilitate morphosyntactic learning in an optimal way.

5.4.3 Comparison of Experiments 1 and 2

To compare the results of participants exposed to the blocked and random training methods statistically, a 2x4 ANOVA was carried out on each set of test data, examining the influence of training method (blocked, random) and training mode (prose, rhyme, melody, rhyme-and-melody). Post hoc test results will be reported as significant if both the by-participants and the by-items analyses reached p-values < .05. Familiarity (novel-old, novel-new) was not

included as an independent variable in the analyses of participants' performance in the generalization tests, since all prior results had shown that, if anything, it affected overall accuracy but it did not interact with training group. As with Experiments 1 and 2, the training modes will be referred to as P (prose, no rhyme, no melody), R (rhyme, no melody), M (no rhyme, melody) and R&M (rhyme and melody).

Old items (Task 4, Test 1, Test 2). The results of the production and comprehension tests carried out with old items are presented in Table 4 and Table 5.

Production tests. The analyses of the test results of Task 4 (random-order sentence production at the end of the last training session) yielded a main effect of training method, $F1(1, 88) = 50.51$, $p < .001$; $F2(1, 71) = 190.76$, $p < .001$, with the blocked groups (Experiment 1) performing significantly better than the random groups (Experiment 2). Furthermore, the results revealed a main effect of training mode, $F1(3, 88) = 5.28$, $p = .002$; $F2(3, 213) = 41.3$, $p < .001$. Post hoc tests yielded significantly better results in the R&M mode in comparison to the P mode, $p = .001$. The interaction of training method (blocked, random) and training mode (P, R, M, R&M) was significant in the by-items analyses only, $F1(3, 88) = 1.44$, $p = .237$; $F2(3, 213) = 19.18$, $p < .001$.

When repeating the same task at the beginning of the test session (Test 1, random-order sentence production) the participants of the blocked training method still attained significantly better results than the participants of the random training method, $F1(1, 88) = 40.33$, $p < .001$; $F2(1, 35) = 111.68$, $p < .001$. Once again, the analyses yielded a main effect of training mode, $F1(3,88) = 6.99$, $p < .001$; $F2(3, 105) = 31.23$, $p < .001$. Post hoc tests revealed significantly better results of the modes M and R&M in comparison to mode P. Critically, there was a significant interaction of training method and training mode for this test, $F1(3,88) = 2.74$, $p = .048$; $F2(3, 105) = 13.62$, $p < .001$. As discussed in the separate analyses of each experiment, training mode affected the performance of the participants in the blocked training method in a more consistent fashion than those trained in the random training method. Recall that significant effects in Experiment 1 were obtainable in both the analyses by-participants and by-items whereas, in Experiment 2, they were apparent by the by-items analyses only.

In the next step, the type of errors participants made when producing sentences during training and test (blocked/random-order sentence production in Sessions 2 and 3, random-order sentence production at the end of the last training session (Task 4) and in the test session (Test 1)) were explored. Four different error types were analyzed: a) an error in which the case-like category of the marker was correct, i.e. a marker from the correct locative category was produced, but the marker was from the wrong gender-like category (gender error), b) an error in which the marker was from the correct gender-like category but the case-like category was incorrect (case error), c) an error in which both the gender- and the case-like categories were incorrect (gender-case error) and d) an error that could be either a gender error, a case error or a gender-case error (not classifiable).

Table 6 summarizes the total number of each error type for each task, collapsed across experiments and training modes. Clearly, across sessions, participants mostly made gender errors, whereas problems with choosing the right case appeared at a lesser rate (see Table 6). Gender errors were dominant during all sessions and did not vanish towards the later sessions, indicating that, as anticipated, choosing the correct gender-like category in the production tasks constituted the most prominent problem for the learners. Therefore, all following analyses focus on gender errors.

Table 6

Experiments 1 and 2: Number of Error Types Produced in Session 2 (Blocked/Random-Order Sentence Production), Session 3 (Blocked/Random-Order Sentence Production), Task 4 (Random-Order Sentence Production at the End of the Last Training Session) and Test 1 (Random-Order Sentence Production in the Test Session)

	Total Number of Items	Error Type			
		Gender Error	Case Error	Gender-Case Error	Not Classifiable
Session 2	6912	2127	133	96	1215
Session 3	6912	1365	113	82	851
Task 4	6912	1089	93	73	666
Test 1	3456	683	45	35	358

Figure 3 depicts the percentage of gender errors out of all responses for all eight groups in each task. As apparent by Figure 3, differences of training method increased from Session 2 to Session 3. Prior to analyzing the influence of training method (blocked, random) and mode (P, R, M, R&M) on the proportion of gender errors, the data was arcsine transformed, looking at each task separately.

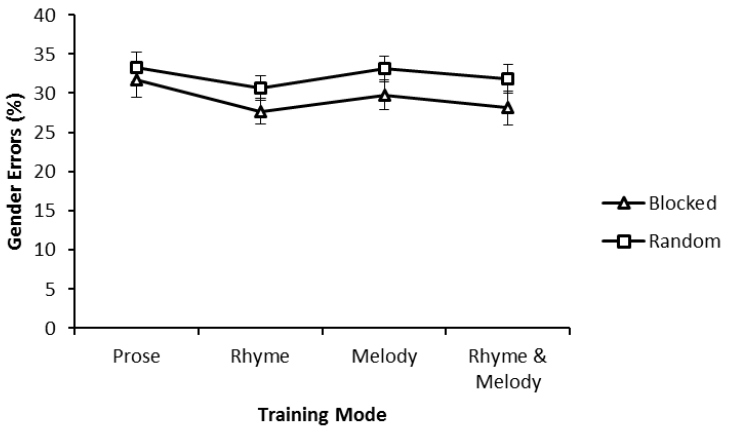


Figure 3: Session 2

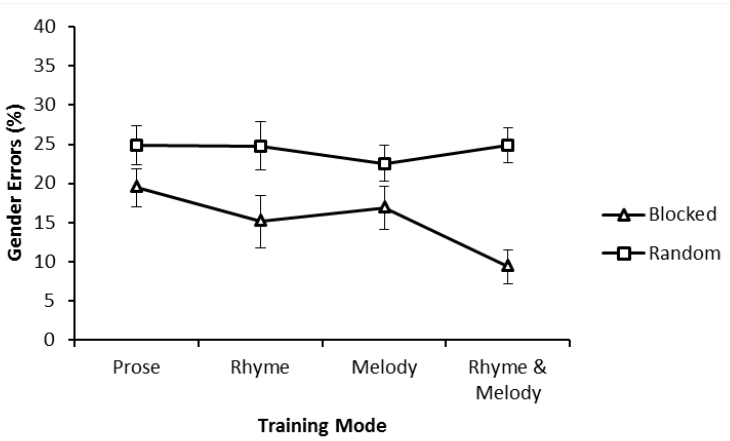


Figure 3: Session 3

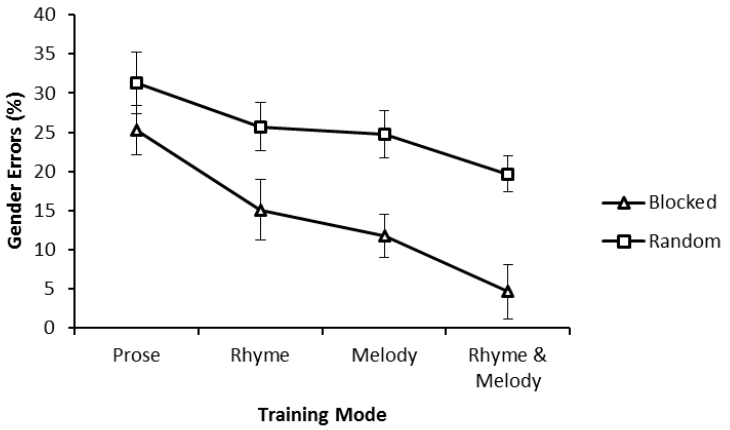


Figure 3: Task 4

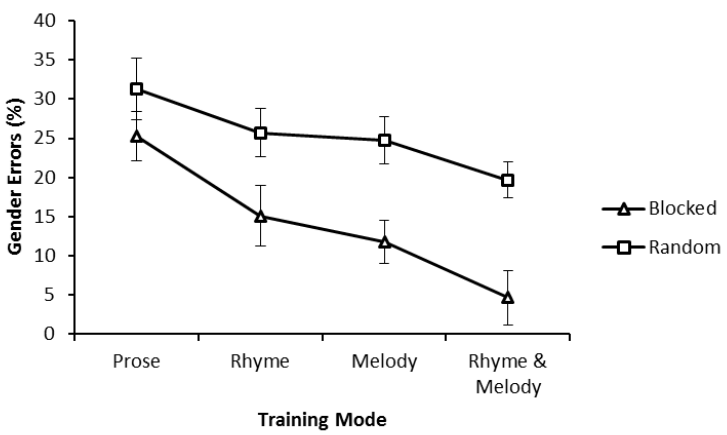


Figure 3: Test 1

Figure 3. Experiments 1 and 2: Mean percentages of gender errors of all responses produced in Session 2 (blocked/random-order sentence production), Session 3 (blocked/random-order sentence production), Task 4 (random-order sentence production at the end of the last training session) and Test 1 (random-order sentence production in the test session).

Table 7 summarizes the results of the statistical analyses of the percentages of gender errors made during the tasks. There was a main effect of training method for all tasks (see Table 7), with participants trained with the random training method producing significantly more gender errors in comparison to those trained with the blocked training method. The advantage of the blocked training method over the random training method was statistically significant from the start, and it substantially intensified from sessions 2 to 3, yielding *F*-values in session 3 that were five times larger than those seen in session 2.

When considering the factor training mode, the results showed a significant effect for Session 2 and Session 3 in the by-items analyses only and a significant effect for Task 4 and Test 1 in both the by-participants and by-items analyses (see Table 7). Thus, the effect of training mode was more obvious for the later sessions, suggesting that differences of training mode manifested themself over time. For Task 4, post hoc tests revealed that participants of the mode P made significantly more gender errors in comparison to those trained in the mode R&M. For Test 1, post hoc tests showed that significantly more gender errors were made in the mode P than in the modes M and R&M. The interaction between training method and mode reached significance in the by-items

analyses for Session 3 only, and it approached significance in Task 4 (see Table 7). Critically, the interaction was not significant in Test 1, suggesting that the long-term effects of training mode were overall similar in both training methods.

Table 7

Experiments 1 and 2: Results of the Statistical Analyses of Gender Errors Observed in Session 2 (Blocked/Random-Order Sentence Production), Session 3 (Blocked/Random-Order Sentence Production), Task 4 (Random-Order Sentence Production at the End of the Last Training Session) and Test 1 (Random-Order Sentence Production in the Test Session)

	df(F1)	*F1*	*p(F1)*	*df(F2)*	*F2*	*p(F2)*
			Training Method (blocked, random)			
Session 2	1, 88	4.88*	.030	1, 71	3.35(*)	.072
Session 3	1, 88	23.88***	.001	1, 71	34.65***	.001
Task 4	1, 88	26.58***	.001	1, 71	57.43***	.001
Test 1	1, 88	23.53***	.001	1, 35	23.31***	.001
			Training Mode (P, R, M, RM)			
Session 2	3, 88	1.28	.287	3, 213	3.48*	0.17
Session 3	3, 88	1.21	.310	3, 213	6.59***	.001
Task 4	3, 88	4.13**	.009	3, 213	22.37***	.001
Test 1	3, 88	8.41***	.001	3, 105	19.45***	.001
			Training Method x Training Mode			
Session 2	3, 88	.116	.951	3, 213	.281	.839
Session 3	3, 88	1.57	.202	3, 213	8.14***	.001
Task 4	3, 88	.328	.805	3, 213	2.58(*)	.055
Test 1	3, 88	.677	.568	3, 105	1.30	.278

Note. (*) $p < .075$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Taken together, the effect of training mode manifested itself over time, with participants of the modes M and R&M outperforming those of the mode P at the end of the study. Training with the blocked method led to significantly less gender errors in comparison to training with the random method. This difference between the two training methods was significant from the start and increased

substantially from task to task, with mode R&M yielding the fastest-growing difference between the random and blocked training methods.[4]

Comprehension test. The 2x4 ANOVA of the levels of marker accuracy in the forced-choice grammaticality judgment test (Test 2) yielded a main effect of training method, $F1(1,88) = 52.698$, $p < .001$; $F2(1, 35) = 28.35$, $p = .009$, with participants of the blocked groups performing significantly better than those of the random groups. There were also a main effect of mode, $F1(3,88) = 11.01$, $p < .001$; $F2(3, 105) = 18.55$, $p < .001$ and a significant interaction between training method and mode, $F1(3,88) = 3.624$, $p = .016$; $F2(3, 105) = 9.44$, $p < .001$. Post hoc tests showed that levels of performance in the mode M turned out to be significantly better in comparison to those achieved in the mode P. Additionally, significantly better results in the mode R&M in comparison to the modes P and R were observed. As shown in the separate analyses of Experiments 1 and 2, effects of mode showed a more consistent pattern within the blocked training method (see Experiment 1).

To summarize, the results of the tests working with old items revealed that blocked learning significantly exceeded random learning. In all three tests (Task 4, Test 1, Test 2), participants of the blocked training method outperformed participants of the random training method. Furthermore, best results could be attained when both rhyme and melody were added to the blocked training method. Taken together, the results of the tests with old items demonstrate that gender-like category induction is best achievable when participants are trained in a blocked rhyme-and-melody method.

Generalization tests (Tests 3 and 4).

Production test. The test results of Test 3 (novel+hint sentence production) are summarized in Tables 4 and 5. The 2x4 ANOVA yielded a main

[4] In order to assess the interaction of time, training mode and training method, an overall analysis of the proportions of gender errors observed over the course of Session 2 to Session 4 was carried out. For reasons to do with the way the data were coded, I was unable to carry out these analyses by items. The by-participants analysis yielded significant main effects of time, $F1(3, 264) = 102.18$, $p < .001$, training method $F1(1, 88) = 28.9$, $p < .001$, and training mode, $F1(3,88) = 4.84$, $p = .004$, with mode P yielding significantly more gender errors than mode R&M. There were significant interactions of time and training method, $F1(3, 264) = 8.48$, $p < .001$, and time and training mode, $F1(9, 264) = 4.19$, $p < .001$, but there was no three-way interaction ($F1 < 1$). This suggests that in terms of the learning trajectories, the effects of method and mode were not superadditive but additive, with the best results seen in training mode R&M when employed with a blocked training method (cf. Figure 3).

effect of training method, *F1*(1, 88) = 158.51, *p* < .001, *F2*(1, 23) = 258.78, *p* < .001 with significantly better results of the blocked training method compared to the random training method. Furthermore, the results revealed a main effect of training mode, *F1*(3, 88) = 3.21, *p* = .027; *F2*(3, 69) = 22.42, *p* < .001. Although no significant effects were shown in the post hoc tests, the mode R&M attained marginally significantly better results than the mode P (both *p*s < .054). The interaction of training method and training mode was significant in the analyses by items (*F2*(3, 69) = 3.73, *p* = .015) but not by participants (*F1*(3, 88) = .520, *p* = .670). As seen in the separate post-hoc tests (by items) in each experiment, Experiments 1 and 2 differed slightly with respect to the mode effect: While in Experiment 1, participants of Groups B-M and B-R&M outperformed the participants of Groups B-P and B-R, Experiment 2 only yielded significant differences between Group R-P on the one hand and Groups R-M and R-R&M on the other.

Comprehension test. Tables 4 and 5 present the results for Test 4 (novel forced-choice grammaticality judgment). The 2x4 ANOVA yielded a main effect of training method, *F1*(1, 88) = 172.59, *p* < .001; *F2*(1, 23) = 71.01, *p* < .001 with significantly better results in the blocked groups than in the random groups. Additionally, the analyses revealed a main effect of training mode, *F1*(3, 88) = 4.5, *p* = .006; *F2*(3, 69) = 5.54, *p* = .002. Post hoc tests showed that the mode R&M yielded significantly better results than the mode P. Comparable to Test 3, the interaction of training method and mode reached significance in the by-items analyses only, *F1*(3, 88) = .927, *p* = .431; *F2*(3, 69) = 2.86, *p* = .043, confirming the differences found in the by-items post hoc tests for Experiments 1 and 2. While in both experiments, mode R&M yielded the best results and mode P yielded the worst results, the experiments differed with respect to the second best mode, M (Experiment 1) and R (Experiment 2), respectively.

In summary, both generalization tests showed significantly better levels of performance in the blocked training method compared to those achieved in the random training method. This result suggests that a blocked presentation facilitates gender-like category induction. Furthermore, the analyses yielded better results of the mode R&M in comparison to the mode P, apparent by a trend in the novel+hint sentence production test (Test 3) and a significant effect in the novel forced-choice grammaticality judgment test (Test 4). Similar to the results of the tests with old items, these results indicate substantial advantages of

being trained in a) a blocked training method and b) a rhyme-and-melody mode. In both tests with old items (Task 4, Test 1, Test 2) and generalization tests (Tests 3 and 4), best levels of performance could be achieved when participants were trained in the blocked rhyme-and-melody method (Group B-R&M).

Questionnaire. To examine background information, such as explicit knowledge of the paradigm or information about L2 experience, participants filled out a questionnaire at the end of the study. A copy of the questionnaire is attached in the Appendix (see Appendix B). The investigation of L2 experience will be assessed in section 5.5.3 (Comparison across languages). The main focus of the questionnaire was the question of whether participants had gained explicit knowledge of the artificial system and, if so, to what extent. Since explicit instructions were missing during all sessions, this issue was particularly interesting. Participants were asked to indicate whether they noticed that the sentences followed certain rules and, if so, to write down those rules.

Table 8 shows the accuracy of paradigm knowledge. Answers were analyzed concerning four different types of paradigm knowledge: a) correct and complete knowledge of the gender-case-like system, meaning that participants could describe which markers were used in connection to which action (e.g. “ino, ano and ono go along with walking towards a creature”) and which sets of markers belonged together (e.g. “ino goes along with gla and sla”) (cf. knowledge of full paradigm in Table 8), b) complete knowledge of the case-like (locative) markers only, i.e. participants were able to describe which markers occurred in connection with which actions but could not describe which set of markers belonged together (locative), c) incomplete knowledge of the case-like (locative) markers, i.e. participants could describe an extract of the case-like paradigm (e.g. “ino goes along with walking towards a creature”) (partial locative) and d) no knowledge of the system at all.

Very clearly, most participants of the blocked training method provided a completely correct description of the system (see Table 8). In contrast, most participants of the random training method showed explicit knowledge (in whole or partial) of the locative paradigm only and failed to establish the whole system. Three participants, all of whom had been trained with the random method, showed no evidence of any knowledge of the marker system. A chi-square test revealed a significant relationship of training method (blocked vs. random) and

type of knowledge (full, locative, partial locative) ($\chi^2_{df=2}$ = 45.15, p < .001), confirming the connection between training method and type of paradigm knowledge. The participants without any knowledge of the marker system were not included in this test, because in their group, the expected frequency was smaller than 5.

Table 8

Type of Gender-Case-Like Paradigm Knowledge for Groups of the Blocked Training Method (Experiment 1) and the Random Training Method (Experiment 2) as Indicated in the Questionnaire

	Type of Paradigm Knowledge			
Training Mode	Full	Locative	Partial Locative	None
	Blocked Training Method (Experiment 1)			
P	8	3	1	0
R	8	3	1	0
M	11	1	0	0
R&M	11	1	0	0
	38	8	2	0
	Random Training Method (Experiment 2)			
P	0	3	7	2
R	0	6	6	0
M	2	6	4	0
R&M	3	5	3	1
	5	20	20	3

To evaluate the impact of paradigm knowledge on the results gained in the random-order sentence production test (Test 1), a chi-square test was carried out. The numbers of correct and incorrect markers produced in Test 1, with respect to the type of paradigm knowledge, are presented in Table 9. As might be expected, the type of knowledge a participant was able to establish during training (full, locative, partial locative, none) had a significant influence on the number of correct markers produced in Test 1 (see Table 9; $\chi^2_{df=3} = 270.68$, $p < .001$).

Table 9

Experiments 1 and 2: Number of Correct and Incorrect Markers Produced in Test 1 with Regard to the Type of Paradigm Knowledge

Markers	Type of Paradigm Knowledge			
	Full	Locative	Partial Locative	None
Correct	1248	599	412	43
Incorrect	300	409	380	65

Concluding, the answers given in the questionnaire indicate that most participants of the blocked training method gained complete knowledge of the morphosyntactic paradigm, whereas most participants of the random training method acquired the locative rules only. Moreover, the results support the assumption that better levels of performance, as seen within the blocked training method, are associated with a more complex and complete knowledge of the system.

In summary, the results replicate Taraban's (2004) findings, demonstrating that the acquisition of an artificial grammar system can be facilitated if the paradigm is presented in a blocked method. Furthermore, they show that training with a combination of rhyming and melodic materials, like in childrens' songs, led to best levels of performance. In all examined tests, best results could be attained when combining the three components input structuring/blocking, rhyme, and melody.

5.5 Experiment 3: Structured vs. unstructured gender-like category induction in English native speakers

Experiment 1 and 2 tested German native speakers. These speakers have acquired a complex gender-case system during their first language acquisition. The domain of the artificial gender-case-like system used in the present study expressed locative conditions and did not refer to thematic roles such as gender and case. However, one might claim that mastering the systematicity of such a complex system could facilitate the acquisition of a similar system in other languages. Thus, the next step is to include speakers of a first language without a gender-case system. By comparing abilities of native speakers of a language with and without a gender-case system, one can examine if previous knowledge of a specifically structured grammatical system influences gender-like category induction. Therefore, Experiment 3 tested English native speakers with selected methods of the training study.

5.5.1 Method

Participants. 50 participants from the University of Illinois at Urbana-Champaign, USA, participated in this experiment in exchange for payment. All participants were English native speakers and aged between 18 and 30 years.

Materials and design. The materials were identical to those used in Experiments 1 and 2 except that all pseudowords and sentences were pronounced in standard American English. The focus of the present work is to examine the role of rhyme, melody, and a structured/blocked input presentation. Thus, the four groups of the blocked training method B-P (prose, no rhyme, no melody), B-R (rhyme, no melody), B-M (no rhyme, melody) B-R&M (rhyme and melody) were trained in the present experiment. As in Experiment 1, the first B in each group's name refers to the blocked presentation of the training material. In order to compare the blocked with the random training method, one random group was included as a baseline group. As witnessed by the results of Experiment 2, the random prose group (Group R-P) attained the worst results in all tests. Therefore, this group was additionally trained, resulting in five groups in total.

Procedure. Participants were randomly assigned to one of the five groups B-P, B-R, B-M, B-R&M or R-P. The procedure was identical to the procedure for Experiments 1 and 2, with participants being exposed to their respective training method (blocked or random) and training mode (prose, rhyme, melody or rhyme-and-melody).

5.5.2 Results and discussion

The data analysis was similar to the analysis for Experiments 1 and 2. As in Experiments 1 and 2, the usage of marker accuracy was analyzed. Sentences including the wrong use of an object or verb occurred rarely (205 out of 6600 cases, totaling 3.11% of the data) and were not excluded from further analyses. It is noteworthy that, although four of the five groups of the present experiment were trained in a blocked fashion (as in Experiment 1), the percentage of incorrect objects and verbs was more similar to the percentage seen in Experiment 2 (4.18%) than of that in Experiment 1 (0.24%).

One-sample t-tests revealed that all five groups scored significantly greater than chance in all analyzed tests ($p > .05$) (see Footnote 2), indicating that language learning appeared in all groups. This result differed from the results reported by Brooks et al. (1993) who showed that the random group did not exceed chance level.

Again, the data was arcsine-transformed. To investigate the differences between the five groups, one-factor ANOVAs with the factor training group (B-P, B-R, B-M, B-R&M, R-P) were performed on the sets of data with old items (Task 4, Test 1, Test 2). For the generalization tests (Test 3, Test 4), the independent variable familiarity was included in order to examine the impact of novel-old vs. novel-new nouns on levels of performance. Hence, for Tests 3 and 4, ANOVAs were carried out with group and familiarity as independent variables. For all analyses, the independent variable training group was a between-participants variable for the participant analyses (F1) and a between-items variable for the item analyses (F2). Post hoc analyses (Tukey HSD) will be reported as significant if the analyses showed p-values < .05.

Table 10

Experiment 3 with Native English Speakers (Blocked/Random Presentation in Training): Percentages of Correct Markers in Task 4 (Random-Order Sentence Production with Old Object Nouns at the End of the Last Training Session), Test 1 (Random-Order Sentence Production with Old Object Nouns), Test 2 (Forced-Choice Grammaticality Judgment with Old Object Nouns), Test 3 (Novel+Hint Sentence Production, Collapsed Over Novel-Old and Novel-New Nouns) and Test 4 (Novel Forced-Choice Grammaticality Judgment, Collapsed Over Novel-Old and Novel-New Nouns)

Training Group	Old Items			Novel- Items	
	Production		Comprehension	Production	Comprehension
	Task 4	Test 1	Test 2	Test 3	Test 4
Blocked Prose	76.53 (19.12)	68.89 (19.98)	76.67 (12.84)	95.00 (15.81)	98.33 (4.03)
Blocked Rhyme	74.03 (13.35)	62.78 (13.11)	76.11 (11.28)	98.33 (5.27)	97.50 (5.62)
Blocked Melody	73.61 (14.25)	61.11 (14.7)	75.56 (9.24)	95.00 (15.81)	98.33 (4.03)
Blocked Rhyme & Melody	80.69 (16.49)	69.72 (18.68)	81.94 (10.08)	95.00 (10.54)	97.50 (4.03)
Random Prose	61.53 (16.13)	59.17 (17.57)	69.17 (9.11)	58.33 (18.00)	73.33 (17.15)

Note. Standard deviations are given in parentheses.

Old items.

Production tests. As mentioned above, the random-order sentence production test was carried out at the end of the last training session (Task 4) and in the test session (Test 1), so that short-term versus long-term learning could be considered. Table 10 contains the percentages of marker accuracy produced in those tests.

In Task 4, the ANOVAs yielded a main effect of training group in the by-items analyses only, $F1(4, 45) = 2.07$, $p = .101$; $F2(4, 284) = 13.41$, $p < .001$. The post hoc tests by items showed that the random group (Group R-P) attained significantly worse results in comparison to the four blocked groups B-P, B-R, B-

M, and B-R&M. When considering differences within the blocked training groups, post hoc tests by items revealed significantly better results of Group B-R&M in comparison to Groups B-R and B-M. Recall that training group is a between-participants but within-items factor. Effects may therefore be observable in the by-items analyses only.

In Test 1, again, the main effect of training group was significant across items but not across participants, $F1(4, 45) = 1.05$, $p = .394$; $F2(4, 140) = 3.55$, $p = .009$. Post hoc tests by items revealed significantly better results of Groups B-P and B-R&M in comparison to those attained in Groups R-P and B-M. A descriptive analysis by participants showed that two out of 50 participants attained 100% correct responses and that these two participants were trained in the Groups B-P and B-R&M.

A further analysis was conducted on the type of errors participants made in the blocked/random-order sentence production tasks that were carried out during training and test. As in the German data analysis, errors were classified as gender errors, case errors, gender-case errors or as not classifiable (see Comparison of Experiments 1 and 2, section 5.4.3). Table 11 presents the total number of each error type for each task.

Table 11

Experiment 3: Number of Error Types Produced in Session 2 (Blocked/Random-Order Sentence Production), Session 3 (Blocked/Random-Order Sentence Production), Task 4 (Random-Order Sentence Production at the End of the Last Training Session) and Test 1 (Random-Order Sentence Production in the Test Session)

	Total Number of Items	Error Type			
		Gender Error	Case Error	Gender-Case Error	Not Classifiable
Session 2	3600	1159	37	51	684
Session 3	3600	767	25	13	551
Task 4	3600	565	20	19	358
Test 1	1800	382	18	11	231

Clearly, participants mainly made gender errors during both training and test, suggesting that gender assignment constituted the major difficulty. The proportion of gender errors produced during the different sessions was analyzed. Table 12 summarizes the percentages of errors produced out of all the responses for each task. As can be seen in Table 12, the frequency of gender errors decreased over time, with the overall smallest value given at the end of the last training session (Task 4).

Table 12

Experiment 3: Percentages of Gender Errors Produced in Session 2 (Blocked/Random-Order Sentence Production), Session 3 (Blocked/Random-Order Sentence Production), Task 4 (Random-Order Sentence Production at the End of the Last Training Session) and Test 1 (Random-Order Sentence Production in the Test Session)

Traininig Group	Session 2	Session 3	Task 4	Test 1
Blocked Prose	35.14 (10.52)	18.19 (9.13)	14.17 (12.66)	19.17 (13.45)
Blocked Rhyme	33.89 (5.16)	22.08 (5.98)	15.56 (8.46)	23.61 (9.47)
Blocked Melody	29.72 (7.97)	21.11 (8.46)	14.17 (9.66)	22.22 (9.71)
Blocked Rhyme & Melody	31.25 (6.39)	17.64 (10.54)	12.36 (11.70)	19.72 (15.24)
Random Prose	30.97 (7.61)	27.5 (10.18)	22.22 (11.71)	21.39 (11.19)

Note. Standard deviations are given in parentheses.

In order to analyze the influence of training group on the proportion of gender errors, the data was arcsine-transformed and one-factor ANOVAs were conducted for each task. The results are summarized in Table 13. As depicted by Table 13, the main effect of group reached significance in the by-items analyses of Session 3 and Task 4, and it approached significance in Session 2. For both Session 3 and Task 4, post hoc tests by items revealed that the random Group

R-P made significantly more gender errors than all four blocked groups and that Group B-R made significantly more gender errors in comparison to Group B-R&M. Critically, there was no significant effect in Test 1, suggesting similar long-term effects in all five training groups.

Table 13

Experiment 3: Results of the Statistical Analyses of Gender Errors Observed in Session 2 (Blocked/Random-Order Sentence Production), Session 3 (Blocked/Random-Order Sentence Production), Task 4 (Random-Order Sentence Production at the End of the Last Training Session) and Test 1 (Random-Order Sentence Production in the Test Session)

	dfs	*F1*	*p(F1)*	*dfs*	*F2*	*p(F2)*
	Effects of Group (BP, BR, BM, BRM, RP)					
Session 2	4, 45	.852	.500	4, 284	2.27(*)	.062
Session 3	4, 45	1.93	.122	4, 284	7.84***	.001
Task 4	4, 45	1.22	.316	4, 284	11.54***	.001
Test 1	4, 45	.212	.930	4, 140	.811	.520

Note. (*) p < .075, * p < .05, ** p < .01, *** p < .001.

Comprehension test. Table 10 presents the percentage of correct responses in the forced-choice grammaticality judgment test (Test 2). The analyses yielded a main effect of training group, which was significant by items but not by participants, $F1(4,45) = 1.72$, $p = .162$; $F2(4, 140) = 5.09$, $p = .001$. Post hoc tests by items showed that Group R-P attained significantly worse results than Groups B-P, B-R and B-R&M. Moreover, the results of Group B-R&M were significantly better than those achieved in the Groups B-R and B-M. Similar to the results obtained in Test 1, the two participants that achieved 100% correct results were trained in the Groups B-P or B-R&M, reflecting the results found in the by-items analyses.

Overall, the results of the training material tests showed that Group B-R&M attained the best and Group R-P the worst results in all training material tests. The analyses revealed an advantage of the blocked training method in comparison to the random training method. Although these results were apparent by the by-items analyses alone; they still highlight that a blocked training method

as well as the rhyme-and-melody mode can strengthen the learning of trained material in both production and comprehension.

Generalization tests.

Production test. Table 10 shows the percentages of marker accuracy in the novel+hint sentence production test (Test3), collapsed across novel-old and novel-new object nouns. Previous analyses had shown that familiarity of novel nouns did not affect participants' performance (both *F*s < 1). However, the interaction of familiarity (novel-old, novel-new) and training group reached significance in the by-items analyses, $F1(4, 45) = 1.41$, $p = .246$; $F2(4, 88) = 2.51$, $p = .047$. Participants of Group B-R&M were more accurate when producing sentences containing unfamiliar (98%) than familiar (92%) object nouns, while the reverse pattern applied to Group R-P (60% correct answers for familiar object nouns vs. 57% correct answers for unfamiliar object nouns). The remaining Groups B-P, B-R, and B-M showed similar patterns of results for both familiar (novel-old) and unfamiliar (novel-new) object nouns.

As apparent by Table 10, all blocked groups achieved excellent results in Test 3 and clearly outperformed Group R-P. Accordingly, the ANOVAs yielded a significant effect of training group, $F1(4, 45) = 16.92$, $p < .001$; $F2(4,88) = 77.58$, $p < .001$, with significantly worse results of Group R-P in comparison to all four blocked groups. Additionally, post hoc tests by items revealed significantly better results of Group B-R in comparison to Groups B-P, B-M, and B-R&M.

Comprehension test. The results of the novel forced-choice grammaticality judgment test (Test 4) are summarized in Table 10. Participants were more accurate when working with unfamiliar object nouns (94%) than with familiar object nouns (92%), $F1(1,45) = 6.31$, $p = .016$; $F2(1,22) = 2$, $p = .172$. Critically, familiarity (novel-old, novel-new) did not interact with training group (both *F*s < .543, $p > .705$).

As apparent by Table 10, the blocked groups attained 97,5% or more correct whereas Group R-P attained worse results. The analyses revealed a main effect of training group, $F1(4, 45) = 15.07$, $p < .001$; $F2(4,88) = 23.74$, $p < .001$. Similar to the results found in the novel+hint sentence production test (Test 3), post hoc tests showed that Group R-P attained significantly worse results than all four blocked groups.

Taken together, the findings revealed significantly better results in the blocked groups in comparison to the prose group regarding both generalization tests. Participants trained in the blocked training method were able to use and generalize the learned rules. Comparable results within the blocked groups could be seen in both production and comprehension generalization, whereas participants of the random Group R-P attained better results in the comprehension test (see Table 10), suggesting that participants of the blocked training method were confident in both production and comprehension of the paradigm rules to a similar extent. The present results indicate that an excellent learning of the underlying grammar system had taken place in the blocked training method.

Questionnaire. As mentioned previously, participants filled out a questionnaire in which they were asked to indicate whether they noticed that the language possessed certain rules, and if so, to write down the rules. The answers were analyzed and assigned to one of the following four types of paradigm knowledge: a) full knowledge of the paradigm, b) locative knowledge, c) partial locative knowledge and d) no knowledge (see Comparison of Experiments 1 and 2, section 5.4.3).

Table 14 summarizes the type of paradigm knowledge for each of the five groups. Clearly, participants of the blocked training method mostly provided explicit knowledge of the full system whereas the participants of the random training method showed (in whole or partial) locative knowledge. This pattern is similar to the one found in the German data (see Comparison of Experiments 1 and 2, section 5.4.3). A chi-square test yielded a significant relationship of training group (B-P, B-R, B-M, B-R&M, R-P) and type of paradigm knowledge (full, locative, partial locative) ($\chi^2_{df=8} = 31.08$, $p < .001$), corroborating the results indicated in Table 14. The one participant who showed no paradigm knowledge was not included in this test. Table 15 shows the relation between type of paradigm knowledge and marker accuracy in the random-order sentence production test (Test 1). The type of paradigm knowledge (full, locative, partial locative, none) influenced the maker accuracy in Test 1 significantly (see Table 15; $\chi^2_{df=3} = 32.17$, $p < .001$).

Table 14

Questionnaire Results: Type of Gender-Case-Like Paradigm Knowledge for the Five Groups Trained in Experiment 3

	Type of Paradigm Knowledge			
Training Group	Full	Locative	Partial Locative	None
B-P	8	2	0	0
B-R	9	1	0	0
B-M	9	1	0	0
B-R&M	7	3	0	0
R-P	0	5	4	1
	33	12	4	1

Table 15

Experiment 3: Number of Correct and Incorrect Markers Produced in Test 1

Markers	Type of Paradigm Knowledge			
	Full	Locative	Partial Locative	None
Correct	787	288	71	12
Incorrect	401	144	73	24

The results of the questionnaire indicate that the English native speakers were able to extract and verbalize the underlying paradigm rules. The distribution of paradigm knowledge is similar to that seen in the German data, with most participants of the blocked training method providing full paradigm knowledge and participants of the random training method providing locative paradigm knowledge. Comparable to the results seen in Experiments 1 and 2, a better and

more specific knowledge of the paradigm, as seen in the blocked training method, is accompanied by a better performance in Test 1.

In summary, Experiment 3 tested the impact of training method (blocked vs. random) and mode (prose, rhyme, melody, rhyme and melody) on the acquisition of gender-like categories in English native speakers. As noted above, in comparison to the German native speakers tested in Experiments 1 and 2, these speakers had not had access to a complex gender-case system during their first language acquisition. The results from Experiment 3 demonstrate two basic findings.

First, they suggest that gender-like category induction can be facilitated when presented in a blocked training method, even if the participants have not been exposed to a comparable system during their first language acquisition. This result is apparent by the generalization tests and the by-items analyses regarding the tests with old items. The one random group (Group R-P) attained worst results in all tests regarding old items, which is consistent with the findings of Experiments 1 and 2. Also, out of all five training groups, Group R-P made the most gender errors throughout the later training units. Furthermore, the results seen in the questionnaire suggest that being trained in the blocked method is associated with a more complex and complete explicit paradigm knowledge.

Significant differences of training method could be manifested more evidently within the generalization tests in comparison to the differences found in the tests with old items. Also, similar to the German data, participants of the blocked training method attained better results in the generalization tests than in the tests with old items (see Table 10). Why does marker accuracy as well as the effect of training method differ between the tests with old items and the generalization tests? As mentioned previously (see Experiment 1, section 5.3.2), participants had to use the correct locative rules and to assign the appropriate gender-like subclass to the object noun when working with old items. In contrast, the generalization tests required them to use and generalize the correct locative rules while the gender-like subclass of the new nouns was given and did not need to be accessed. Thus, the tests with old items might be harder to complete than the generalization tests. The findings mentioned above are consistent with this assumption.

The second basic finding indicates that rhyme and melody can promote gender-like category indcution in the trained material. An advantage of the Group

B-R&M in comparison to other blocked groups was shown in all three tests with old items, demonstrating a similar pattern to the results seen in Experiments 1 and 2. However, it is noteworthy to state that the present results were found in the by-items analyses only. This suggests that the improvement of Group B-R&M was caused by certain participants only, but that those participants improved across all items. When considering the amount of gender errors made during training and test, Group B-R&M showed least errors during the later training units. In contrast to Experiments 1 and 2, Group B-R&M failed to attain best results in the generalization tests (see Table 10). Moreover, unlike in Experiment 1, Group B-P did not achieve the worst results when comparing results within the blocked groups. Rather, the analyses showed a clear advantage of Group B-P in Test 1 and no significant differences between the Groups B-P and B-R&M regarding all tests. Taken together, the present results highlight the advantage of the rhyme-and-melody mode once again, however, the findings are not as consistent as in the German data (see Experiment 1, section 5.3.2).

Overall, the results found in Experiment 3 highlight the facilitation of gender-like category induction in English native speakers through a) a blocked training method and b) a rhyme-and-melody mode, and are therefore in line with the results found in Experiments 1 and 2.

5.5.3 Comparison across languages

Participants of Experiments 1 and 2 were native speakers of German, who have already acquired a complex gender-case system during their first language acquisition (see Table 1). The systematicity of the artificial system used in the present work (see Table 3) was based upon an extract of the German gender-case system, i.e. the overlap in the markers within and across the gender-like categories was comparable to the overlap given in the German extract. Although the chosen artificial paradigm referred to locative conditions, instead of thematic roles and was thus different from the German gender-case system, the German participants' (implicit) knowledge about structure and usage of such a complex system might have facilitated the acquisition of the system in Experiments 1 and 2. Hence, the purpose of Experiment 3 was to find out to what extent native speakers of a language without a gender-case system acquire the artificial gender-like subclasses of the artificial language.

Therefore, Experiment 3 tested native speakers of English with selected methods of Experiments 1 and 2. Recall that the first two experiments examined the impact of training method (blocked in Experiment 1 versus random in Experiment 2) and mode (prose, rhyme, melody, rhyme-and-melody in both Experiments 1 and 2) on the acquisition of the artificial subclasses. Thus, eight groups of German native speakers were trained in Experiments 1 and 2. Due to the effects found in the results of Experiments 1 and 2 (see Comparison of Experiments 1 and 2, section 5.4.3), Experiment 3 was designed to include all four modes of the blocked training method (B-P, B-R, B-M, B-R&M) plus the prose mode of the random training method (Group R-P), resulting in five training groups when English native speakers were tested. In the following, the pattern of results seen in German and English native speakers will be compared. First, training and test results of both languages groups will be considered and compared. Thereafter, the questionnaire data will be evaluated.

Old items. When comparing the results between German and English native speakers in the tests working with old items (Task 4, Test 1, Test 2), a similar pattern of results could be observed. In both language groups, Group B-R&M attained the best and Group R-P the worst results concerning both production (Task 4, Test 1) and comprehension (Test 2) tests. Thus, both German and English data underline the advantage of the blocked training method along with the rhyme-and-melody mode.

Yet, the findings between German and English native speakers revealed differences in terms of consistency. While the German data clearly indicate the advantage of the blocked training method (see Comparison of Experiments 1 and 2, section 5.4.3), only a few English native speakers showed effects of training method (see Experiment 3, section 5.5.2). Effects of mode were most consistent in German participants of the blocked training method (Experiment 1), revealing significant results in both the by-participants and by-items analyses. In contrast, the findings regarding German participants of the random training method (Experiment 2) and English native speakers (Experiment 3) both revealed significant differences of training group in the by-items analyses only. This result suggests that the observed group differences were caused by the majority of participants in Experiment 1 and only by a few participants in Experiments 2 and 3.

When considering the type of errors produced in training and test, both German and English data revealed that gender errors constituted the dominant problem (see Comparison of Experiments 1 and 2, section 5.4.3 and Experiment 3, section 5.5.2). Moreover, in both languages, participants of the random training method made significantly more gender errors than participants of the blocked training method, confirming the results described above. However, considerable differences in the improvement over time could be seen between participants of both languages. Whereas performances of German and English native speakers showed a similar level in the first training session (Session 2), German participants showed greater improvements from Session 2 to Session 3 and overall better performance levels from Session 3 onwards in comparison to the English-speaking participants.

To examine the impact of native language (German, English) on participants' marker accuracy, a 2x5 ANOVA was conducted on each set of test data, examining the influence of language (German, English) and training group (B-P, B-R, B-M, B-R&M, R-P). Only those groups who were trained in both the German and English setting (all four blocked groups and the random prose group) were included in these analyses. Overall, in the tests working with old items, German native speakers attained significantly better results than English native speakers (Task 4: $F1(1,100) = 2.59$, $p = .111$; $F2(1, 71) = 15.88$, $p < .001$; Test 1: $F1(1,100) = 8.57$, $p = .004$; $F2(1, 35) = 27.96$, $p < .001$; Test 2: $F1(1, 100) = 12.49$, $p = .001$; $F2(1, 35) = 8.34$, $p = .007$). While the main effect of language reached significance in both the by-participants and the by-items analyses for Tests 1 and 2, it was found in the by-items analyses only for Task 4, suggesting that the differences between German and English native speakers were even greater in long-term learning. The analyses yielded a main effect of training group (Task 4: $F1(4, 100) = 9.88$, $p < .001$; $F2(4, 284) = 72.15$, $p < .001$; Test 1: $F1(4,100) = 7.27$, $p < .001$, $F2(4, 140) = 27.03$; Test 2: $F1(4, 100) = 12.79$, $p < .001$; $F2(4, 140) = 22.36$, $p < .001$), confirming the results seen in the separate analyses of Experiments 1 to 3. Recall that, for both the German and the English data, the best results were achieved by training groups B-R&M and the worst results were achieved by training groups R-P. Critically, there were significant interactions of language and training group (Task 4: $F1(4, 100) = 2.25$, $p = .069$; $F2(4, 284) = 18.53$, $p < .001$; Test 1: $F1(4, 100) = 5.14$, $p = .001$; $F2(4, 140) = 22.49$, $p < .001$; Test 2: $F1(4, 100) = 4.04$, $p = .004$; $F2 (4, 140) = 10.15$, $p <$

.001). As mentioned above, the effects of training method and training mode were more consistent in the German data than in the English data. Note that the interaction effects were significant for Tests 1 and 2 in both the by-participants and the by-items analyses. In contrast, for Task 4, the interaction effect reached significance in the by-items analyses only. This pattern of results suggests that differences of language manifested themself over time.

To sum up, the overall pattern of results for the tests with old items was similar in all three experiments, with Group B-R&M yielding the best and Group R-P the worst results. The effects were more consistent for the German data than for the English data. This finding provides support for the assumption that a blocked training method as well as the combination of rhyme and melody can facilitate gender-like category induction in both speakers of a native language with and without a complex gender-case system. However, for these tests working with old items, German native speakers attained significantly better results in comparison to those seen in English native speakers, demonstrating an advantage of German participants regarding the acquisition and mastering of trained material.

Generalization tests. The generalization tests examined the mastering of the trained paradigm regarding new material. Therefore, new object nouns were introduced. As mentioned above, half of those new nouns had already been worked with during the vocabulary training and vocabulary review but were otherwise excluded from language training (novel-old nouns, i.e. familiar nouns). The other half of the nouns had not been introduced during training and were thus unfamiliar (novel-new nouns) (see also Appendix A).

For both German and English data, familiarity did not play a role in the novel+hint sentence production test (Test 3). The results for the novel forced-choice grammaticality judgement test (Test 4) differed from those found in Test 3, with participants of both language groups yielding higher levels of accuracy when responding to unfamiliar than to familiar nouns. Critically, familiarity did not interact with training group. Thus, with respect to the relation between familiarity and performance, similar results were found in both German and English data.

When comparing results gained in the two generalization tasks (see Experiment 1, section 5.3.2, Experiment 2, section 5.4.2 and Experiment 3, section 5.5.2), a similar pattern was observed for German and English native

speakers. It is striking that both German- and English-speaking participants of the blocked training method attained excellent, comparable results. Those participants yielded levels of accuracy between 92.36 and 100%. In contrast, participants of the random training method attained significantly worse results in both language groups.

As for the effect of mode, differences could be manifested in the by-items analyses only (see Experiment 1, section 5.3.2, Experiment 2, section 5.4.2 and Experiment 3, section 5.5.2). This pattern of results applies to both German and English data. However, the effects differed slightly between both language groups. On the one hand, effects of mode were demonstrated for both Test 3 and Test 4 in the German data but only for Test 3 in the English data. On the other hand, the data gained from German native speakers indicate a clear advantage of the R&M mode in all tests, whereas the English data highlights the R mode in Test 3 as a particularly effective mode.

To compare the generalization performance between German and English native speakers, again, a 2x5 ANOVA was conducted on both sets of test data. The analyses revealed no significant effects of language concerning both generalization tests (Test 3: $F1(1, 100) = .39$, $p = .534$; $F2(1, 23) = .247$, $p = .624$; Test 4: $F1(1, 100) = .473$, $p = .493$; $F2(1, 23) = .126$, $p = .726$), indicating that participants of both language groups were able to use and generalize the gained rules to a similar extent. The analyses yielded a main effect of training mode (Test 3: $F1(4, 100) = 48.05$, $p < .001$; $F2(4, 92) = 102.61$, $p < .001$; Test 4: $F1(4, 100) = 48.73$, $p < .001$; $F2(4, 92) = 48.65$, $p < .001$), confirming the results shown in the separate analyses for both language groups. Furthermore, for Test 3, the by-items analyses revealed an interaction effect of language and training group, $F1(4, 100) = 1.46$, $p = .221$; $F2(4, 92) = 27.25$, $p < .001$). No interaction effect was found for Test 4, $F1(4, 100) = 1.32$, $p = .269$; $F2(4, 92) = 1.4$, $p = .242$. This pattern of results confirms that, as described above, the differences between the German and English generalization data are rather small.

The fact that the results attained by German and English native speakers significantly differed in the tests with old items, but not in the generalization tests, merits further discussion. While German participants outperformed English participants in the tests pertaining to trained material (see above, section *Old items*), no significant differences were found for the tests working with new material (generalization tests). This finding is in line with the assumption that the

generalization tests were easier to perform than the tests with old items (see Experiment 1, section 5.3.2 and Experiment 3, section 5.5.2), most likely on account of the hint given in the generalization tests but not in the tests with old items. Furthermore, it is consistent with the result that participants trained in the blocked groups performed better in the generalization tests than in the tests with old items, yielding levels of accuracy near ceiling in both generalization tests. This pattern of results could be shown for both German and English native speakers.

Overall, the generalization tests underline the role of a blocked training method. The results of both German and English native speakers indicate that the ability to use and generalize the artificial rules can be enhanced when training takes place in a blocked method. Participants of both language groups attained comparable results regarding production and comprehension generalization, suggesting that they acquired and mastered the artificial paradigm to a similar extent. Effects of mode were more consistent in the German data, highlighting the rhyme-and-melody mode regarding both generalization tests for the German participants.

Questionnaire. At the end of the study, participants filled out a questionnaire in which they were asked whether they have noticed certain rules of the artificial languages, and, if so, to write down the rules. Clearly, the majority of participants gained some kind of paradigm knowledge (see Tables 8 and 14). Only three out of 96 German participants and only one out of 50 English-speaking participants gained no paradigm knowledge at all. The comparison of the type of knowledge displayed by participants of the blocked and random training groups yielded parallel patterns for the German and English native speakers. In both language groups, most participants of the blocked training method demonstrated full explicit knowledge about the system, whereas the majority of the participants trained in the random method provided locative knowledge (in whole or partial) and failed to combine the locative rules with the assignment of gender-like categories.

Remarkably, all of the four participants who were unable to detect rules were trained with the random training method. When examining the influence of training method on type of knowledge, both German and English data revealed a significant relationship. Furthermore, a significant relationship between type of

knowledge and marker accuracy in the random-order sentence production test (Test 1) was found for both German and English native speakers.

In addition to the type of knowledge, participants indicated which further foreign languages they had learned during their life. The answers were evaluated with regard to the total number of acquired foreign languages and the number of acquired gender-marking languages, meaning languages that differentiate between two or more gender categories, such as German, Spanish or Russian. In supplementary analyses, I assessed the influence of the number of acquired languages on marker accuracy in Experiments 1 to 3. In order to be able to compare the German and English speaker groups, only those groups that were trained in both the German and English setting, that is, all four blocked groups (Groups B-P, B-R, B-M, and B-R&M) and the random prose group (Group R-P), were considered for these analyses.

Participants had acquired between 0 and 6 foreign languages, with an average value of 2.26 foreign languages per participant. Table 16 summarizes the number of foreign languages and gender-marking languages for German- and English-speaking participants. Note that groups consisted of 12 participants in the experiments conducted in Germany (Experiments 1 and 2), whereas only 10 English native speakers were tested per group (Experiment 3). As apparent by Table 16, the range of foreign languages differed between the language groups, with German participants speaking between 1 and 6 foreign languages and English participants speaking between 0 and 3 foreign languages. On average, German native speakers had acquired 2.95 foreign languages whereas English native speakers had learned 1.44 foreign languages. This difference yielded significance, $t(108) = 8.68$, $p < .001$. As for the number of acquired gender-marking languages, the data indicate a similar pattern (see Table 16), with significantly more gender-marking languages acquired by German than by English native speakers, $t(108) = 9.60$, $p < .001$. On average, the German-speaking participants had learned 2.85 gender-marking languages whereas the English-speaking participants had learned 1.24 gender-marking languages. Given that both groups of participants were undergraduate and graduate students, this considerable difference could relate to the distinction in the school systems in Germany and the United States of America. The German school system requires students to at least study two foreign languages in order to

qualify for college. In contrast, American students mostly study foreign languages on a voluntary basis in form of elective classes.

Table 16

Number of Acquired Foreign Languages and Gender-Marking Languages as Indicated by German and English native speakers

	Number of Foreign Languages						
Native Language	0	1	2	3	4	5	6
German	0	4	19	18	15	3	1
English	1	28	19	2	0	0	0
	1	32	38	20	15	3	1

	Number of Gender-Marking Languages						
	0	1	2	3	4	5	6
Native Language							
German	0	4	22	17	13	4	0
English	3	33	13	1	0	0	0
	3	37	35	18	13	4	0

As apparent by Table 16, the group of English native speakers can be divided into those participants who had studied one or no foreign language (29 out of 50) and those who had studied two or three foreign languages (21 out of 50). To examine whether there was a connection between number of foreign languages and marker accuracy in the random-order sentence production test (Test 1), a chi-square test was conducted. The results revealed a significant relationship of number of foreign languages and marker accuracy in Test 1 ($\chi^2_{df=1}$ = 4.65, p = .031), with better marker accuracy seen in those participants who had studied more foreign languages. When conducting a similar analyses for the German participants, no significant influence of number of acquired foreign languages (one to two foreign languages versus three to six foreign languages, see Table 16) on marker accuracy in Test 1 was found ($\chi^2_{df=1}$ = 2.62, p = .105). This finding points to the conclusion that the number of acquired foreign

languages can facilitate gender-like category induction in an artificial language when the native language lacks a gender-case system.

To summarize, data gained from both German and English native speakers could demonstrate that input structuring (through blocking) can facilitate gender-like category induction, that of which is apparent by all tests. The results of the generalization tests showed a similar pattern for both language groups with participants of the blocked training groups scoring near ceiling and participants of the random prose groups attaining significantly worse results. This suggests that both German- and English-speaking participants were able to use and generalize their paradigm knowledge to a similar extent. Both the German and the English data of the present study replicate Taraban's (2004) finding that structuring the input and thus optimizing syntactic cues can be sufficient for the effective acquisition of artificial gender-like subclasses, even when phonological cues are lacking. Moreover, both the German and English data indicate that being trained in a blocked method is associated with a more complete knowledge of the paradigm.

Another outcome shown by both language groups is the finding that the combination of rhyme and melody can support gender-like category induction when added to a blocked training method. The effect is apparent by the tests working with old items and is notably stronger and more consistent within the German data. When comparing levels of performance regarding sentence production in training and test, the findings indicate that assigning the correct gender-like category was the dominant problem in both languages groups. However, faster improvements and better results in the later sessions could be found for the German participants. This observation is consistent with the overall significantly better results of German native speakers in the tests with old items.

Finally, the data shows that German-speaking participants had acquired more foreign languages in total and more gender-marking languages than the English-speaking participants. Additionally, those English-speaking participants who had studied a greater amount of foreign languages showed better marker accuracy in the test session. This leads to the conclusion that the knowledge of more (gender-marking) languages is linked to better long-term learning effects.

6 General discussion

6.1 Summary and discussion of the main findings

The purpose of the present study was to find out under what conditions artificial gender-like subclasses can best be acquired. To this end, I trained different groups of participants with an artificial grammar system that contained three gender-like categories. Additionally, there were three case-like categories, which referred to the locative conditions to, by and from. That is, participants watched one artificial creature, “Tika”, walking towards (to-locative), around (by-locative) or away from (from-locative) the gender-marking-relevant object noun creature, respectively.

None of the artificial object nouns contained phonological noun cues. As discussed previously (see section 2), the validity of noun cues is limited for the majority of early acquired German monomorphemic nouns. Therefore, in the present study, the distribution of the locative markers was the only hint as to the gender-like category of the noun. In order to use the markers accurately, participants had to pick up these syntactic cues, i.e. find out about the order and grouping of markers regarding the three locative conditions to, by and from. Additionally, they had to acquire and memorize the correct gender-like category for each object noun. The systematicity of the marker allocation (see Table 3) was based on an extract of the German gender-case system (cf. Table 1), with the artificial paradigm containing the same overlapping number of markers as in the German extract.

In three experiments, I examined in what ways three forms of input presentation, namely input structuring/blocking, rhyme, and melody, facilitate gender-like category induction. Experiments 1 and 2 tested German native speakers whereas Experiment 3 was conducted with English native speakers. By testing two language groups, I aimed to find out whether the acquisition of a highly complex gender-case system in the first language, as given for the German native speakers, affects the acquisition of the artificial system.

In Experiment 1, the role of rhyming and melodic presentation forms along with a structured input presentation was investigated. Thus, the syntactic cues were optimized by presenting all sentences that belonged to one object noun as

a group. In other words, all input during language training was presented in a blocked fashion in this experiment. As described in section 4, a blocked input typically occurs in various forms of language plays, such as poems or songs. Additionally, language play naturally entails rhyming and melodic structures. The objective of Experiment 1 was to investigate the effect of the three major presentation forms of language play as well as their interaction. Participants, all German native speakers, were trained with the blocked input in one of the four training modes prose (B-P), rhyme (B-R), melody (B-M) or rhyme-and-melody (B-R&M). The major results are summarized in the following.

The accuracy of all groups significantly exceeded chance for all analyzed tests (production and comprehension tests in both tests with old items and generalization tests). This finding indicates that gender-like category learning had taken place in all training groups. As for the differences between the training groups, the blocked rhyme-and-melody group (Group B-R&M) attained the best results in all tests. The results of the tests with old items revealed that levels of performances increased from Group B-P to Group B-R&M. Furthermore, they demonstrated a clear advantage for Group B-R&M, with significantly better results than Group B-P in the production tests and significantly better results than all three other groups regarding the comprehension test.

The results yielded by the generalization tests are not as consistent as those seen in the tests with old items; however, they point towards the same direction. In both the production and comprehension generalization test, participants of all four groups achieved excellent results, with values ranging between 92% and 100% correct answers (see Table 4). This indicates that a blocked input presentation is sufficient for gender-like category learning. Yet, as apparent by the by-items analyses, the results still revealed an advantage of being trained with melody (Group B-M) or the combination of rhyme and melody (Group B-R&M).

Overall, Experiment 1 shows that gender-like category induction can be facilitated by structured, blocked input, even in the absence of noun cues. This finding clearly replicates Taraban's (2004) results. The effect increased when the blocked input was accompanied by both rhyme and melody. These outcomes were shown for the learning of a relatively complex system that contained three gender-like categories and marker overlapping, both of which are seen in the German gender-case system.

The objective of Experiment 2 was to examine the role of rhyme and melody when added to a random input presentation. In comparison to Experiment 1, which featured input comparable to that of language plays, this random input is rather typical for everyday language interactions. Since I was able to replicate Taraban's (2004) findings with the results shown by Experiment 1, the major intent of Experiment 2 was to find out to what extent rhyme and melody alone would facilitate gender-like category induction. Therefore, in this experiment, all German-speaking participants were trained with a random training method, presenting all training sentences in a random, unstructured order. Again, participants were trained in one of the four training modes prose (R-P), rhyme (R-R), melody (R-M) and rhyme-and-melody (R-R&M). Similar to the results seen in Experiment 1, participants' accuracy was significantly greater than chance regarding all tests, suggesting that learning occurred when training took place in the random training method.

Moreover, the findings of both tests with old items and generalization tests are in line with those seen in Experiment 1, demonstrating worst results for training in the prose mode (Group R-P) and best results for training in the combination of rhyme and melody (Group R-R&M). However, significant effects of training group were mostly shown by the by-items analyses only. This indicates that the group differences were caused by only a few participants, whereas the differences seen in Experiment 1 were due to the majority of participants. Thus, despite the fact that the effect of training mode (prose, rhyme, melody, rhyme-and-melody) was reduced when training occurred in the random training, the overall pattern was still similar to the results of Experiment 1.

To sum up thus far, syntactic cues are sufficient for the acquisition of artificial gender-like categories, even when the markers are presented in a random fashion. However, as shown by the results of Experiment 1, syntactic cues can be optimized by presenting the input in a structured, blocked way. Finally, the combination of rhyme and melody facilitates learning, especially when added to a blocked input.

When comparing results of the blocked training method (Experiment 1) with the random training method (Experiment 2) in an overall analysis, three main findings were apparent. First, training with the blocked method led to significantly better results than training with the random method, replicating Taraban's (2004) finding that gender-like category induction is supported by input structuring.

Second, the results revealed a clear advantage for the rhyme-and-melody mode regarding both tests with old items and generalization tests. This result underlines the findings reported for the separate analyses of Experiments 1 and 2. Third, participants mainly made gender errors when producing sentences, suggesting that choosing the correct gender-like category was most difficult. An analysis conducted on the gender errors made in the production tasks during both training and test demonstrated that the advantage of the blocked training method was significant from the start, whereas the advantage of the modes melody and rhyme-and-melody became more obvious and stronger during the later sessions of the study. Taken together, the results showed that gender-like category induction was most successful when learning took place in a combination of the three presentation forms input structuring/blocking, rhyme and melody.

Given that the findings were demonstrated for German native speakers and that those speakers had acquired a highly complex gender-case system during their first language acquisition, a fundamental question of mine was whether this knowledge might have helped participants to learn and generalize the artificial system. In order to find out about the potential value of former experience to gender and case, Experiment 3 tested English native speakers with selected methods of Experiments 1 and 2. More precisely, all four blocked training groups (B-P, B-R, B-M, B-R&M) and one random prose training group (R-P) were included in this experiment.

The results of all training groups significantly exceeded chance concerning all tests, replicating the findings reported in Experiments 1 and 2 and thus indicating that morphosyntactic learning had taken place in all groups of both German- and English-speaking participants. Furthermore, the results of Experiment 3 revealed a clear advantage for the blocked training method in comparison to the random training method. However, it is noteworthy that these latter results were apparent by the by-items analyses only. With respect to the effects of training mode, the results of the English-speaking participants were not as consistent as the German data (cf. Experiments 1 and 2): While the German data clearly demonstrated an advantage of the rhyme-and-melody mode regarding all tests, the English data replicated this effect for the tests with old items only.

As for the error analyses, gender assignment constituted the prominent problem for the English-speaking participants, which is in agreement with the German data. Again, training with the blocked method led to significantly less gender errors during the training sessions. However, in contrast to the results of Experiments 1 and 2, this finding was apparent by the by-items analyses only. Moreover, there was no effect of group in the test session, demonstrating similar levels of performance for all five training groups at the end of the experiment. These findings indicate a less consistent pattern in comparison to the results seen in the German data.

To summarize the results of Experiment 3, the English data suggests that artificial gender-like category induction can be facilitated by a blocked training method, even when the participants had not been exposed to a gender-case system during their first language acquisition. Second, the findings indicate that the combination of rhyme and melody can further support this facilitation, which is, in general, in line with the German data. However, the results are less consistent than those of Experiments 1 and 2, showing advantages of the blocked rhyme-and-melody group (Group B-R&M) in the by-items analyses of the tests with old items only. Overall, the German and English data are in accordance, highlighting the value of input structuring/blocking and the combination of rhyme and melody when artificial gender-like subclasses are acquired.

A comparison of the results attained by both language groups revealed differences between German- and English-speaking participants for the tests with old items: The German-speaking participants showed greater improvements in the beginning of the study and overall better results than the English-speaking participants. As for the generalization tests, there were no significant differences between the language groups, indicating that both German and English native speakers acquired the underlying morphosyntactic paradigm to a similar extent. Even though the findings of both language groups differed in terms of consistency, they showed a similar pattern of results. The overall main results of the present study can be summarized in two findings: First, a blocked input presentation leads to significantly better results in comparison to a random input presentation. Second, the combination of a rhyming and melodic presentation can facilitate gender-like category induction, especially when added to the

blocked training method. This pattern was shown for both German and English data.

With respect to the first findings, the present study supports the assumption that drawing learners' attention towards syntactic cues can facilitate the acquisition of gender-like categories when phonological cues are lacking (cf. Taraban, 2004). While the present study replicates the results reported by Taraban (2004) in this regard, it is noteworthy that it differed from Taraban's (2004) study in several aspects, such as the presentation mode (oral presentation in the present study vs. written presentation in Taraban's (2004) study) or the number of gender-like categories (three gender-like categories in the present study vs. two gender-like categories in Taraban's (2004) study). Also, within the blocked training method, Taraban (2004) randomized the presentation order of sentences belonging to one noun. That is, although sentences belonging to one noun were grouped together, the order of sentences differed within each group. In contrast, sentences of one group were always presented in the same order (to-locative, by-locative, from-locative) in the present study. This order was necessary due to the musical structure implemented in the modes melody and rhyme-and-melody. A certain melody was associated with each locative sentence, resulting in a chorus-like melody when all three sentences were presented as a group. Despite these differences, the crucial outcome of the present study is similar to that of Taraban's experiments (see Taraban, Experiments 3 and 4). The data indicate that using syntactic cues can be sufficient for the acquisition of gender-like subclasses when a highly structured input is given.

In contrast to preceding studies (e.g. Brooks et al., 1993; Frigo & MacDonald, 1998; Taraban, 2004), the present gender-case-like system was based upon an extract of a naturally existing gender-case system. Participants of the present study were exposed to a morphosyntactic system expressing three gender-like categories in three case-like categories. Additionally, the artificial system contained marker overlapping across the nine syntactic functions. Finally, the nouns did not exhibit any phonological cues as to their gender. All three features are characteristic for the German gender-case system and therefore make it a highly complex system, which is especially hard to acquire for second language learners. In the present study, all groups demonstrated learning, suggesting that, despite the complexity of the system, learning took place.

Critically, learning increased when a well-structured input was given. Other studies have confirmed the important role of input complexity, demonstrating that a higher input variability leads to better grammar learning (e.g. Brooks et al., 2006; Gomez, 2002).

With regard to the second finding, the results overall demonstrated an advantage for the rhyme-and-melody mode. Hence, it can be assumed that the combination of a rhyming and melodic input presentation can be particularly useful in gender-like category learning. This result is in line with the findings discussed above (cf. section 3). A large body of literature exists that demonstrates the potential value of music for language processing and language acquisition processes. Beyond the numerous findings providing evidence for the direct link between music and language (e.g. Saffran et al., 1996b; Saffran et al., 1999; Schön et al., 2010; Steinbeis & Koelsch, 2008), music has been shown to promote processes that play a role in morphosyntactic language learning, e.g. the detection of phonological boundaries or the segmentation of words (e.g. Schön et al., 2008). Rhyming abilities have been shown to be linked to better language skills at various levels, such as the development of reading and writing abilities (e.g. Wood & Terrell, 1998a) or articulation proficiency (e.g. Mann & Foy, 2007). Even though no studies have evaluated the impact of rhyming abilities on morphosyntactic learning, there is growing evidence of the beneficial role of rhyming abilities on language proficiency and language learning.

As discussed in chapter 4, rhyme, melody, or a combination of both are typically present in many forms of language play. Furthermore, language play naturally presents language in a structured way. The positive impacts of input structuring, rhyme and melody when presented in playful language have been pointed out in section 4. The “Generative Textproduktion” (GT) is a didactic concept that combines the three presentation forms for promoting German grammar learning (e.g. G. Belke, 2012) and has been addressed in detail in section 4.2. The implications of the present findings for promoting the acquisition of complex morphosyntactic systems will be discussed in section 6.3.

A further fundamental finding of the present study relates to the fact that many participants were able to verbalize their acquired morphosyntactic knowledge. As shown by the questionnaire results, the majority of English and German native speakers were able to explain, at least partly, the underlying regularities of the artificial system. The main result was that increased accuracy,

as observed within the blocked groups, was associated with a better and more complete knowledge of the artificial system. A more precise examination of this relationship revealed that better knowledge was linked to better results in the random-order sentence production test (Test 1), which worked with material that participants had previously been trained with. This test was conducted during the test session and required participants to assign the correct gender-like category to a noun as well as to correctly use the case-like locative marker paradigm. The results showed that accuracy in using the case-like paradigm plus accessing previously learned information as to the gender-like category of the particular noun is related to better knowledge about the artificial system. This finding is in line with the view proposed by van den Bos and Poletiek (2010), who assume that, in implicit learning situations, “memorizing the individual exemplars is facilitated by acquiring knowledge of the underlying structure” (p. 139).

What does the fact that participants can verbalize knowledge about the acquired system tell us about the learning processes that had occurred during the training phase? Recall that participants of the present study did not receive any explicit instructions as to the artificial language, methods or aims of the study. Rather, prior to the study, they were informed that no further information about the language would be given before the completion of the study. They were only instructed with respect to the conduction of the separate tasks. For instance, when completing Test 1, they were asked to watch the action and to try to produce the associated sentence. Thus, the participants were exposed to a learning situation that lacked explicit instructions concerning the critical features of the language.

Williams (2005) proposes that verbal report can be seen as the best predictor as for the question whether a stimulus has been noticed. As mentioned above, most participants were able to identify some kind of knowledge regarding the artificial system. Participants of the blocked training method mainly showed knowledge of the complete system whereas participants of the random training method mostly demonstrated partial knowledge. Thus, input structuring seems to support the establishment of knowledge about the complete gender-case-like marker system. This result is consistent with the assumption that knowledge gained in artificial grammar learning can be partly consciously identified, even when explicit instructions are missing (e.g. De Jong, 2005; Pothos, 2007; van den Bos & Poletiek, 2010). It is further in accordance with the view that this

knowledge is linked to better performances (e.g. van den Bos & Poletiek, 2010; Williams, 2005).

Since explicit instructions and explanations about the structure of the artificial language were absent in the present study, the learning situation was rather implicit, at least in the beginning of the training phase. Indeed, various studies support the idea that implicit learning might lead to explicit knowledge. For instance, Mathews et al. (1989) demonstrated that implicitly acquired knowledge about artificial grammars can be, to a certain extent, consciously reflected and verbalized. In different experiments, the authors trained participants with finite-state grammars while explicit instructions were either present or absent. The results revealed that participants of the implicit learning situations were able to gain and verbalize abstract knowledge about the artificial system. The authors thus suggest that different learning processes can be intertwined when complex tasks are acquired. The assumption that artificial grammar learning might combine an implicit and an explicit component was confirmed by other studies (see Pothos, 2007 (pp. 229-232) for an overview).

In a recent study, van den Bos and Poletiek (2010) further examined the relationship between artificial grammar learning and explicit knowledge building. In two artificial grammar learning experiments, they trained participants with materials that either contained a highly salient (Experiment 1) or a less salient (Experiment 2) feature that was predictive with respect to the structure and the distinction of two artificial grammars. Participants were visually exposed to two different artificial grammars, which were referred to as the left and the right grammar concerning their visual orientation on the screen. The predictive feature distinguishing between the left and the right grammar consisted in the initial letter of the strings (highly salient feature, Experiment 1) or the second letter (less salient feature, Experiment 2). In the test phase, participants were instructed to indicate whether strings occurring in the middle of the screen belonged to the right or the left group. Additionally, in order to test explicit knowledge about the grammars, participants wrote down whether they had noticed any differences between the left and the right group and, if so, to define the differences.

The results showed that in both experiments, some participants provided explicit knowledge about the structures of the grammars. Furthermore, similar to the results seen in the present study, a link between explicit knowledge and levels of performance was shown. Crucially, the number of participants

demonstrating explicit knowledge was significantly greater when training took place with a highly salient feature in comparison to training with a less salient feature, indicating the important role of salience when explicit knowledge is built (van den Bos & Poletiek, 2010).

This finding parallels the results of the present study, which demonstrate that input optimization leads to better knowledge about the artificial morphosyntactic system. By grouping together sentences that belonged to one object noun and thus presenting the sentences in a blocked fashion (cf. blocked training method in Experiments 1 and 3), it is likely that the morphosyntactic system presented in the input was more salient than in the random presentation (cf. random training method in Experiments 2 and 3).

In agreement with the findings of the present study, Taraban (2004) also reported that accuracy in the generalization test was connected with explicit knowledge of the gender-case-like system. In the present study, participants who provided better and more complete knowledge were predominantly trained with the blocked method. Those participants attained overall better results in both the tests with old items and the generalization tests.

A positive link between knowledge about the morphosyntactic paradigm and language proficiency has not only been reported for artificial grammar learning, but also in terms of second language learning. Brooks and Kempe (2013) trained participants with an extract of the Russian gender-case system without giving explicit instructions as to the regularities of the paradigm. None of the participants had had prior experience with Russian or any Slavic or Baltic language. The results revealed that the ability to describe the gender-case system at the end of the study was strongly associated with successful learning. Those participants who demonstrated explicit knowledge about the system were able to use and generalize the case-marking regularities best (Brooks & Kempe, 2013).

Taken together, these findings suggest that becoming aware of a trained morphosyntactic system leads to better proficiency regarding this system in both artificial grammar learning and second language learning. Seger (1994) states that the development of a cognitive task that solely trains implicit learning processes is rather difficult. With respect to the link between implicit learning situations and explicit knowledge, she mentions the possibility of explicit learning developing subsequently to implicit knowledge building. When exposed to an

implicit learning situation, as given in the present study, participants might first notice a pattern. In the next step, they may begin to consciously search for regularities, which, finally, might lead to explicit knowledge of the system. In the present study, the majority of participants consciously searched for regularities when they were first exposed to the artificial language: They developed a feeling that the language was based upon certain rules. This was apparent by the fact that they asked questions as to the artificial system or structure of the sentences, even though they had initially been told that no further information about the language will be given until the study is completed. Beyond the observation that the majority of the participants asked about the system, some participants even verbalized their effort to find out about the rules while carrying out a task.

As reported by the results of the study conducted by Brooks et al. (1993), children were barely able to provide knowledge about the artificial morpho-syntactic system while adult participants were able to describe the system to a greater extent. This might indicate that children do not search for the underlying morphosyntactic rules, or at least to a lesser extent. Future studies will have to show if children are able to gain explicit knowledge about morphosyntactic systems and, if so, whether this knowledge is associated with levels of performance similar to the results described above.

Overall, research has shown that implicit learning can lead to a certain amount of explicit knowledge (e.g. De Jong, 2005; van den Bos & Poletiek, 2010). De Jong (2005) further assumes that the development of explicit knowledge might be mediated through implicit knowledge. The present thesis does not address the question whether the learning process occurred implicitly or explicitly. However, the results show that explicit knowledge has been built despite a lack of explicit instructions and that input structuring might support this process.

Finally, the question of how language experience affects gender-like category induction will be further considered. In the present study, German (Experiments 1 and 2) and English (Experiment 3) native speakers were tested in order to find out about influences of prior language experience. As shown by the results of the questionnaire, the German-speaking participants had acquired a greater amount of foreign languages in general and more languages that differentiate between two or more gender categories (referred to as gender-marking languages). As noted previously, the German-speaking participants

achieved significantly better results than the English-speaking participants regarding the tests with old items. Since those tests required the participants to bring together the case-like locative maker system with the assignment of the correct gender-like category, there are strong reasons to assume that these tests were more complex than the generalization tests. This assumption is further underlined by the fact that differences between the language groups were most obvious in these tests.

Why is prior language experience linked to better results in the more complex tests? Several preceding studies have shown that language experience can positively affect language performance. The study conducted by Kempe and Brooks (2008) was already described and discussed in section 2.1. The authors trained participants with a miniature language that contained an extract of the Russian gender-case system. The participants had not acquired Russian or other Slavic language experiences prior to the study. At the end of the study, participants filled out a questionnaire that included questions as to their extent of foreign language knowledge. The results showed that knowledge of gender-marking languages, such as Spanish or Italian, had a positive effect on the test performance. Spanish and Italian were two of the languages that most participants had studied during their life. Therefore, it is noteworthy to say that Spanish, Italian and Russian share certain similarities, for example the ending -a for feminine nouns in the nominative case. This similarity regarding phonological noun cues might have supported the positive impact of prior L2 experience. Overall, Kempe and Brooks (2008) found that prior experience to gender-marking languages facilitates morphosyntactic learning. In a recent study, the authors replicated this finding, demonstrating that the acquisition of gender-marking languages was positively linked to better performance concerning learning of gender agreement of Russian (Brooks & Kempe, 2013). Again, most participants of this study had studied languages that contained parallels to Russian gender marking regarding phonological noun cues. Crucially, awareness of the Russian gender-case system did not influence this transfer process, indicating that the positive effect of prior knowledge about gender-marking languages for greater levels of performance might proceed implicitly.

The impact of former experiences with gender-marking languages on artificial and natural language learning is supported by several related studies (e.g. Sabourin, Stowe, & de Haan, 2006; Williams, 2005). For instance, Williams

(2005) conducted two experiments with a miniature language comprising an artificial gender-case-like system. In two experiments, participants were exposed to English sentences containing artificial determiners for the object noun. Participants knew that the determiners changed regarding the distance of the object relative to the subject. That is, objects were either near or far from the subjects. However, the participants were not told that the choice of artificial determiners did also depend upon the animacy (living, nonliving) of the object. Thus, participants were exposed to a learning situation that, at least partly, lacked explicit instructions about gender-like category induction.

As shown by the results, knowledge of gender-marking languages was related to increased generalization performance concerning the animacy manipulation, i.e. the gender-like category of the object. Furthermore, the fact that most participants were not aware of the animacy distinction but still scored above chance indicates that prior knowledge of gender-marking languages might facilitate implicit learning processes. Similar to the results of the present study, participants who had acquired a gender-marking L1 also spoke more gender-marking languages in total. With respect to the question of why experience with gender-marking languages leads to better grammar learning, Williams (2005) assumes that general grammar knowledge might be a relevant factor. He points out that none of the participants had acquired languages that categorize words as to their animacy. Thus, it is likely that the result was due to a general familiarization with the systematicity of morphosyntactic systems, which might facilitate the acquisition of a comparable system in a different domain (Williams, 2005).

The results of the present study appear to be consistent with this assumption. Given that the artificial system used in the present experiments comprised locative conditions, it was different from the German gender-case system concerning their domains. However, the systematicity of both the German and the artificial system was similar in terms of the proportion of overlapping markers and the absence of phonological noun cues (cf. Tables 1 and 3). Given that the German-speaking participants attained better results in the tests with old items, one can assume that prior experience to the systematicity of gender-case systems might be linked to better levels of performance in the more complex tasks. However, it remains unclear whether this effect was based on their L1 experience with a gender-marking language or on their greater experience to

foreign (gender-marking) languages in general. Unsurprisingly, the German-speaking participants had also studied more foreign languages in total, including gender-marking languages. Thus, I cannot distinguish whether the transfer effect was due to knowledge of a) their L1, b) gender-marking languages in total or c) foreign languages in general.

The present study highlights that greater knowledge of other languages, as seen in the German-speaking participants, is associated with higher levels of performance regarding familiar test material. Interestingly, as described in section 5.5.3, the impact of prior language experience on the results in Test 1 differed somewhat between German- and English-speaking participants when considering accuracy differences within both language groups: While the amount of foreign languages was linked to test accuracy for the English-speaking participants, no association was found within the German-speaking participants. This finding might indicate that additional L2 experience leads to better performance only when the L1 does not include a highly complex gender-case system. In other words, the acquisition of a complex gender-case system in the L1 might have provided a sufficient basis for successful gender-like category learning, so that differences did not manifest within the German-speaking participants. Overall, the present research suggests that prior exposure to gender-marking languages might indeed promote the acquisition of gender-like categories.

To sum up the main findings of the present study, the results indicate that optimizing syntactic cues through input structuring can support the acquisition of artificial gender-like categories. Also, the combination of rhyming and melodic structures was shown to be effective, especially when it was added to a structured input presentation. Best results were attained when all three presentation forms (input structuring/blocking, rhyme, and melody) were combined. In everyday language promotion situations, this combination naturally occurs when language play is used. Thus, the results provide evidence that input presentation characteristic for language play can be effective when promoting complex grammatical paradigms and associated abstract category knowledge. The implications of the study and the relationship with language promotion will be further discussed in section 6.3. Moreover, the results suggest that explicit paradigm knowledge can develop in the absence of explicit instructions and that a better and more complete knowledge is linked to better test accuracy. Finally,

prior language experience was shown to be related to better results, indicating the positive impact of transfer effects, even when the grammar paradigms differ with respect to their domains.

6.2 Limitations of the study

In the following section, four limitations of the present study will be pointed out and discussed.

The first limitation concerns the fact that the duration of sentence presentation in the training modes differed from one another during the training sessions. In the two training modes rhyme (R) and rhyme-and-melody (R&M), every sentence was followed by a rhyme. Hence, within these modes, the presentation duration of each sentence was substantially longer than that of the stimuli in the training modes prose (P) and melody (M). Therefore, a fundamental question is whether this difference in the duration of the stimuli could be an alternative explanation for the pattern of the results. If so, the findings should reveal that the results of participants who were exposed to a rhyming version (training modes R, R&M) differ significantly from those who were trained in a non-rhyming mode (training modes R, M). However, the results refute this prediction. To begin with, participants of the modes R and R&M did differ in their pattern of results in Experiments 1 and 3. Although the presentation durations were similar for those training modes, participants of the mode R&M attained significantly better results overall than those trained in the mode R. Also, in all three experiments, the patterns of both the +rhyme modes (R, R&M) and the -rhyme modes (P, M) were not similar. If the distinct presentation durations alone would have critically influenced language learning, the pattern of results should differ between the +rhyme and the -rhyme modes. However, as discussed in the separate analyses for Experiments 1 to 3, this is not the case. Thus, it is unlikely that the differences in the presentation duration worked as a crucial factor.

Recall that the present study focused on forms of input presentation that work in language play. As described in section 4, rhyming and melodic structures are important features of language play and are combined in many forms of language play. Hence, the speech input of language play naturally differs from everyday language situations, for example with respect to the duration of the linguistic stimuli or concerning the usage of repeated elements (cf. section 4.1).

The aim of the present study was to examine the presentation forms that work in language play, in isolation and in combination, as well as to compare this input with a typical input of an everyday language situation. Therefore, the present study investigated language learning when the three presentation forms input structuring, rhyme and melody occurred exclusively (Groups B-P, R-R and R-M), when they were combined (Groups B-R, B-M, B-R&M, R-R&M) and when all three were missing (Group R-P). Thus, similar to the distinct input given in language promotion situations using language play versus everyday life language situations, the input of the different training modes differed from one another.

The second shortcoming of the present study is the fact that the duration of input presentation given during the training sessions differed between the blocked and the random training method. More precisely, when working through the listen-and-repeat tasks (Task 1), participants of the blocked training method (see Experiments 1 and 3) were exposed to three sentences as a group. After listening, they were instructed to repeat all three sentences. Due to the combination of input structuring and a melodic structure, grouping three sentences was a natural consequence. In contrast, in the random training method (see Experiments 2 and 3), single sentences were presented and repeated.

One possibility in order to make the input presentations of the training methods more comparable would be to adapt the input of the random training method to the grouped, tripartite structure of the blocked training method. That is, the randomly assigned sentences given in the random training method could be presented in groups of three, similar to the input given in the blocked training method. However, due to the random assignment of sentences featured in the random training method, the input would still differ between the two training methods. More precisely, the three sentences grouped together would relate to three different object nouns in the random training method but to one object noun in the blocked training method. Therefore, I assume that a similar duration of input presentation between both training methods is rather hard to implement.

Again, for reasons of modeling language play situations versus everyday language situations, the difference in input presentation might be reasonable. From a different point of view, the verbal memory span required in the random training method was shorter than in the blocked training method. Thus, if the differences in presentation duration worked as a critical factor, it is likely that the

shorter input presentation would have led to strengthened storage processes and overall better learning. Yet, the results revealed a reversed picture, with overall better learning seen in the blocked groups in comparison to the random groups. Thus, despite the drawback of differences in input presentations between the training methods, it is unlikely that these differences had a critical impact on the results.

A third limitation of the study is that a sufficient examination of the rhythmic component is lacking. As described previously, rhythm plays a great role in the input of language play (cf. section 4). It naturally interacts with other features and presentation forms of language play and supports language processing, such as the segmentation of linguistic units. Thus, rhythm might play a central role when language learning is promoted by language play. Indeed, the findings of previous studies indicate a relationship between rhythmic skills and language abilities.

For instance, Thomson and Goswami (2008) have examined the link between written language abilities and rhythmic skills in children. In their study, they tested 48 10-year-old children with different language and rhythm tests. The children were assigned to two different groups, based on their written language abilities. Children of the dyslexic group had been diagnosed with specific reading impairments, while the other children showed a typical reading development. The tests assessed the children's phonological awareness skills, general motor skills, auditory rhythmic processing and receptive and motor rhythm abilities. As for auditory processing, discrimination skills were examined, namely the abilitiy to discriminate tones or tone sequences that differed in rise time (rise time discrimination) as well as the ability to discriminate tone sequences with regard to their difference in duration (duration discrimination), intensity (intensity discrimination) and frequency (frequency discrimination). The rhythmic ability was measured by using one receptive and one motor test. In the receptive test, the children were asked to discriminate pairs of tone sequences with respect to their different durations of their inter-stimulus-interval (tempi discrimination). The motor rhythm test examined the ability to tap the mouse button along with a metronome beat (paced condition) and to continue tapping after the beat has stopped (unpaced condition).

The results revealed that the dyslexic children performed significantly poorer in all tests of auditory rhythmic processing, except for the intensity

discrimination test. Furthermore, a significant group difference was found for the motor rhythm test, with children of the dyslexic group performing significantly worse at tapping in time of the beat and exhibiting a less constant tapping pattern in comparison to the typically developing group. Those group differences were seen in the paced condition only; in contrast, when tapping after the metronome had ceased (unpaced condition), the majority of children of both groups performed poorly. As apparent by further results, general motor skills were not connected with dyslexia and had no influence on the results of the rhythm tests or the association between rhythm and written language abilities (Thomson & Goswami, 2008).

Taken together, the results reported by Thomson and Goswami (2008) indicate that written language abilities are linked to auditory rhythmic processing skills. Furthermore, they suggest that children with dyslexia have greater difficulties to adapt their motor control to an extrinsic rhythm. Overall, the main finding of the study is that motor rhythmic skills in children seem to be strongly associated with written language abilities (Thomson & Goswami, 2008). Further evidence for the link between motor rhythm skills and written language abilities is supported by the finding that this relationship also appears in adults (e.g. Thomson, Fryer, Maltby, & Goswami, 2006; Wolff, 2002).

Additionally, there has been growing evidence for an association between specific language impairment and rhythm difficulties regarding auditory rhythmic timing (Corriveau, Pasquini, & Goswami, 2007) and motor rhythmic skills (Corriveau & Goswami, 2009), supporting the assumption that rhythmic skills and language skills might be related. Similar to the idea of shared neural systems for music and language, as discussed in sections 3.1 and 3.2, Thomson and Goswami (2008) suppose that the close relationship between rhythm and language development might be due to a neural link between both domains. The authors further suppose that expressive rhythmic activities, such as beating the drum while listening to or singing a song, might positively affect language acquisition (Thomson & Goswami, 2008).

This assumption is in line with research discussed in this thesis (cf. sections 3.1 and 3.2) as well as with the results of the present study, demonstrating the critical value of melody for language acquisition. Since rhythm is an important component of melody, it is likely that rhythmic abilities might be intertwined with the supportive role of melody for language acquisition.

Furthermore, as pointed out previously, rhythmic skills were shown to be associated to rhyming abilities (Wood, 2006). In language play, rhythm is naturally linked to rhyming and melodic structures. Thus, the potential value of rhythm and its combination with rhyme and melody should be examined. However, the present study does not consider this in detail. In the three experiments, a specific rhythm naturally co-occurred in the training modes featuring rhyme and melody, but was absent in the prose mode. Based on the idea to model everyday language situations, the prose mode did not exhibit any particular rhythmic features other than those of a natural rhythm occurring in everyday language. In order to consider the role of rhythm more precisely, one possibility is to include an additional training mode in which the items are spoken along with a specific rhythm (prose-and-rhythm mode). Thus, the training mode prose would be divided into two sub-modes, with rhythm being present or absent. In this way, the exclusive examination of rhythm and its role for language learning would be possible.

Finally, the fourth limitation of the present study is the inability to specifically distinguish between the role of different forms of prior language experience. The results revealed that prior language experience had a positive impact on the test accuracy. However, it is unclear whether this effect was due to knowledge of a) an L1 gender-marking language, b) the number of L2 gender-marking languages, or c) the number of L2 languages in general. It is also conceivable that the combination of an L1 gender-marking language and the knowledge of a greater amount of L2 (gender-marking) languages was crucial for better results, as seen in the German-speaking participants. Although the specific relationship between prior language experience and test accuracy cannot be revealed in its entirety, it can be concluded that language experience might facilitate gender-like category induction.

6.3 Implications for promoting gender category induction

The objective of the present thesis is to answer the crucial question of how the acquisition of opaque gender(-like) categories can be facilitated. Many languages mark for grammatical gender and case; however, when comparing different languages, there are great differences as to the transparency of gender marking. In some languages, such as Spanish of Italian, the gender of a noun is

predominantly phonologically marked on the noun itself (e.g. the ending -a for feminine nouns vs. the ending -o for masculine nouns), at least for a large amount of frequently used and early acquired nouns. In contrast, this is not the case for German gender assignment. Due to the restricted reliability of phonological and semantic cues for gender marking (e.g. Maratsos, 1983), the German gender-case system is considered one of the most complex hurdles any learner of German must face (e.g. MacWhinney, Leinbach, Taraban, & McDonald, 1989). Additionally, the German gender-case system contains marker overlapping, which, along with the lack of noun cues, makes it a highly complex morphosyntactic system that might seem opaque and confusing when learning German as a second language. Thus, it is not surprising that children who learn German as their second language typically show difficulties in acquiring this system.

During their language acquisition, children with German as their second language are usually exposed to a less frequent and less rich German input in comparison to the input native speakers receive. Similarly, there are certain groups of children with German as their native language who suffer from an insufficient German input or are not able to fully process complex input, such as children who grow up in a poor language environment or hearing impaired children. Many of the children noted above exhibit difficulties with the acquisition of the German gender-case system (e.g. Rösch, 2003; Szagun, 2004), which is, for instance, apparent by the incorrect use of determiners or inflected adjectives. When acquiring the German gender-case system, gender assignment appears to be the most pressing problem, at least for second language learners (Jeuk, 2007).

Thus, a crucial question is how the acquisition of the German gender categories can be facilitated and what input presentation can be used in order to support children during that process. In light of the foregoing, the focus on syntactic cues is particularly important because the syntactic environment of a noun often provides distinct hints as to its gender, for example through the co-occurrence of the noun, its determiner and an associated pronoun. Since phonological and semantic noun cues are lacking for a wide range of German nouns, the promotion of the German gender acquisition should focus on optimizing syntactic cues. In the present thesis, I discussed and experimentally investigated this optimization. The optimization of syntactic cues, namely the

systematically grouping of units, forms or structures, is typically given in many forms of language play. Additionally, language play combines this structured input presentation with rhyming and melodic structures. Thus, I aimed to find out in what ways these three presentation forms characteristic for language play may facilitate gender-like category induction. In the following, the impact and supportive role of the three presentation forms for language learning will be pointed out.

First, Taraban (2004) demonstrated that input structuring through blocking associated gender-case-like markers facilitates gender-like category induction in the absence of phonological cues. The results of the present study are compatible with this finding, clearly demonstrating the advantage of a blocked compared to a random input presentation when the syntactic environment was the only hint as to the gender-like category of the noun. These findings support the assumption that gender-like category induction can be facilitated by presenting the input in a well-structured fashion. The same should apply to natural languages that, wholly or partly, lack phonological noun cues, such as German. Here, it is likely that the optimization of syntactic cues may represent a crucial learning factor. Thus, systematic promotion of highly complex grammar systems, such as the German gender-case system, should always include input that is presented in a clearly structured fashion.

In speech and language therapy, this concept is well-established, for instance regarding speech therapy with specific language impaired children (e.g. Siegmüller & Kauschke, 2006). One possibility when working with an optimized, structured input is to group together central grammatical markers or, in other words, to provide a blocked presentation. For instance, one could block the determiners of a certain noun in its different cases, which is exactly the presentation I aimed to model in the present experiments. This input enables the learner to experience the co-occurrence of different grammatical forms and might thus facilitate the acquisition of the underlying correlations of paradigms and categories.

Second, there is extensive literature concerning the role of music for language processing and language learning. Music in general and melody in particular have been proven to work as a supportive factor for different levels of language learning. For instance, studies have demonstrated the beneficial value of musical training for verbal memory (e.g. Ho, Cheung, & Chan, 2003). Other

studies have shown the important role of melody for sequence learning (e.g. Schön et al., 2008; Thiessen & Saffran, 2009). Both verbal memory and sequence learning are important for morphosyntactic learning. With these findings as background, one might conclude that music or, to be precise, a melodic presentation of the linguistic input should be applied systematically in speech therapy and language promotion situations. Melody is typically used, consciously or unconsciously, when natural language promotion takes place in first language acquisition. For example, singing childrens' songs repeatedly is an everyday life routine in parent-child interactions or typical kindergarten settings. Singing is also a central component of phonological awareness promotion and different speech therapy programs. This leads to the conclusion that melodic presentations should be more strongly linked to systematic language promotion situations.

Third, the relationship between rhyming abilities and language skills has been discussed. As shown by various studies, rhyming abilities are linked to different areas of language acquisition, such as articulation (Mann & Foy, 2007), phonological perception (Foy & Mann, 2001) and written language acquisition (Wood & Terrell, 1998a). It is well-known that rhyming abilities play a central role in phonological awareness skills, thus demonstrating the link between both rhyme and language in early language acquisition. However, to my knowledge, no study has examined the role of rhyme in grammar acquisition. Therefore, one fundamental question of the present study was whether rhyme has a positive effect on morphosyntactic learning. The results of the present study show that the combination of music and rhyme was more powerful than the usage of one of these presentation forms alone. Thus, gender-like category induction can be promoted by presenting the input along with both rhyming and melodic structures.

The three presentation forms input structuring, melody and rhyme are most naturally intertwined in many forms of language play, such as songs or clapping rhymes. As section 4 demonstrates, language play occurs in everyday life situations and plays a central role in language acquisition. It predominantly occurs in early parent-child interactions or among children and seems to be particularly important during preschool and early school age (Sanches & Kirshenblatt-Gimblett, 1976). Language plays integrate linguistic form with

semantic content and pragmatic use and are therefore an appealing language input for children.

When using language plays, children do not primarily focus on the linguistic form of the text. However, unconsciously, linguistic forms and structures become the focus of attention (G. Belke, 2012). This is accomplished through certain features of language play, such as the repetition and/or the parallelism of language, systematic blocking of certain linguistic structures or the support of linguistic form by rhyme and rhythm. Due to the combination of these features, linguistic units might be perceived and segmented better, the detection and distinction of linguistic structures and paradigms might be facilitated and the memorization of linguistic associations might be strengthened. Thus, the intertwining of the features, with systematic input structuring, rhyme and melody being most dominant, can lead to overall implicit language learning when language plays are used. It is well-known that implicit learning is crucial in early language learning (E. Belke & Belke, 2006). Language play might be an optimal basis for implicit language learning (G. Belke, 2007a), with different forms of language play facilitating implicit learning of different linguistic levels.

The “Generative Textproduktion” (GT), a didactic concept for groups of language learners, systematically uses language play for language promotion, focusing on the acquisition of linguistic structures and paradigms (e.g. G. Belke, 2012). It basically uses songs and other forms of language play to make complex morphosyntactic systems accessible and comprehensible. For instance, GT systematically works with songs as “Der Katzentatzentanz” (see 3a) to 3c) for a variant adapted from the original) in order to promote the acquisition of the German gender-case system. The concept can be applied to oral and written promotion situations and can thus be used in kindergarten and school settings. It is applicable to language promotion and speech therapy settings for groups of learners as well as for the single child.

As discussed above, language play is considered as an optimal basis for implicit learning processes. Given that certain structures and systems of the German grammar are highly complex, such as the German gender-case system, the systematic introduction into those systems as well as their structured promotion are highly recommended. Since obvious noun cues, i.e. clear rules, are missing in the German gender assignment, language promotion should focus

on implicit learning processes when German nouns and their associated gender categories are to be taught.

Germany is an immigration country and multinational groups and classes of children are part of our everyday life. Many children who learn German as their second language show particular difficulties when acquiring German morphosyntax, with the German gender acquisition constituting one of the most prominent problems. Thus, systematic language promotion of German is strongly required. This is exactly what GT does. Here, certain linguistic structures and forms are presented in a clearly structured fashion, along with rhyming and melodic structures. Children enjoy the usage of language play and will therefore be and stay motivated when GT is applied. Furthermore, both children with low and high language proficiency and with different language backgrounds can be promoted as a group, reflecting the reality of German kindergartens and schools. By linking the implicit acquisition of complex morphosyntactic systems with the creative use of language play, GT might be an optimal way to promote grammar skills.

There are numerous teachers who work with GT in their classrooms. Upon inquiring as to their personal experiences with the concept, those teachers often report that a) children enjoy the work and b) the concept does facilitate the detection and long-term learning of linguistic forms and structures. However, up to this point, the empirical evidence of the forms of input presentation used by GT is lacking. To fill in this gap, I conducted the present study. As shown by the results, input structuring was most powerful. Additionally, the combination of rhyme and melody facilitated language learning significantly. Critically, language learning was most successful when all three forms of input presentation were combined. This combination represents the typical input provided by many forms of language play. If we translate this finding to everyday life language promotion situations, concepts using the combination of all three presentation forms, such as GT, should be powerful instruments for the promotion of complex morphosyntactic systems.

To conclude, the present study clearly highlights the relevance of a systematic use of language play for promoting gender category induction. This might be especially useful for languages that do not possess transparent noun cues, such as German, and for the acquisition of highly complex morphosyntactic

systems, which are best achievable when the use of the crucial structures are presented in systematic way.

7 Conclusions

The present thesis aimed to examine the question of which forms of input presentation may promote gender(-like) category induction. It is well-known that many children, such as children who acquire German as a second language, struggle with the German gender acquisition. A systematic inclusion of promoting German grammar skills, such as German gender assignment, into daily kindergarten and school settings is therefore strongly required. The “Generative Textproduktion” is a didactic concept for the systematic promotion of German in heterogeneous groups of children. The concept works with different forms of language play in order to promote language skills, focusing on morphosyntactic structures and paradigms.

Basic features of language play are that linguistic structures are often grouped together (structured/blocked input) and that the linguistic input is usually accompanied by rhyming and/or melodic structures. Prior research has demonstrated the advantage of music and rhyme for language processing and language acquisition. Furthermore, Taraban (2004) has shown that a structured input presentation facilitates gender-like category induction. Accordingly, the present study investigated in what ways the presentation forms input structuring, rhyme, and melody affect gender-like category induction.

The main findings can be summarized as follows: The greatest positive effect on gender-like category learning was attained through structuring the input, so that sentences belonging to one noun were presented as a group. Additionally, the combination of rhyme and melody led to better accuracy, especially when presented along with structured input. At the end of the study, both German- and English-speaking participants demonstrated similar results when using the regularities of the underlying paradigm. In contrast, the German-speaking participants outperformed the English-speaking participants in the more complex tests, which required the assignment of the correct gender-like category as well as the application of the underlying paradigm regularities. Although no explicit instructions were given, the majority of participants gained, at least partially, explicit knowledge about the artificial system. Better and more complete knowledge was associated with higher accuracy during test. Finally, greater

language experience, as seen for the German-speaking participants, was linked to better levels of performance.

There are several issues that need to be addressed in future studies. One would be to examine the rhythmical component in more detail. Rhythm naturally accompanies rhyming and melodic structures, however, I did not focus on rhythm within this study. Another issue worthy of evaluation would be the complexity of the grammatical system itself. Since the artificial system was based upon a limited extract (3 gender-like categories x 3 case-like categories) of the German gender-case system, the next step would be to find out if similar results can be revealed when the system is made more complex, for instance by introducing number (singular, plural) to the paradigm. Finally, further studies will have to examine whether similar results are obtainable when children are tested. A study by Hudson Kam and Newport (2005) suggests that children tend to overgeneralize grammatical regularities more frequently than adults. One hypothesis following from this is that when learning the present language, children are more likely to overgeneralize the overlapping (and hence more frequent) markers than adults are. If so, it is conceivable that the extent to which such overgeneralizations occur is dependent on the type of training method and mode used.

In summary, the results suggest that optimizing syntactic cues through input structuring can support the acquisition of artificial gender-like subclasses. Furthermore, the combination of rhyme and melody was shown to be effective, especially when it was added to structured input. For both German- and English-speaking participants, best results could be attained when all three forms of input presentation (input structuring, rhyme, and melody) were combined. In everyday language promotion situations, this combination naturally occurs when language play is used. The present results provide evidence that the combination of the three major presentation forms of language play can be effective when promoting the acquisition of complex morphosyntactic systems. Concluding, the present thesis strongly indicates that language promotion can benefit from language play when applied in a well-structured way. It highlights the relevance of concepts that systematically use language play, such as the “Generative Textproduktion”.

8 German summary

Der Erwerb einer Erst- oder Zweitsprache umfasst viele verschiedene Bereiche und Lernprozesse auf allen linguistischen Ebenen. Eine zentrale Erwerbsaufgabe stellt hierbei der Erwerb morphosyntaktischer Strukturen, Regularitäten und Paradigmen dar. Morphosyntaktische Fähigkeiten erlauben es einem Sprecher beispielsweise, über Sätze oder Texte hinweg grammatische Bezüge herzustellen, und sind aus diesem Grund essentiell für komplexere Sprachverwendung. Im Deutschen stellt der Erwerb des Genus-Kasus-Systems einen Kernbereich der morphosyntaktischen Entwicklung dar. Für viele Kinder ist der Erwerb dieses Systems eine große Schwierigkeit, was beispielsweise durch die Verwendung falscher Artikel oder Adjektivflexionen ersichtlich wird. Der Erwerb des deutschen Genus-Kasus-Systems bereitet vor allem Kindern, die Deutsch als Zweitsprache lernen oder Kindern, die in einem spracharmen Umfeld aufwachsen, enorme Schwierigkeiten. Dabei stellt der Bereich der Genuszuweisung für Kinder mit Deutsch als Zweitsprache die größte Erwerbsaufgabe dar (Jeuk, 2007). Die Tatsache, dass 25% aller in Deutschland aufwachsenden Kinder das Deutsche nicht als Erst-, sondern als Zweitsprache erwerben (Autorengruppe Bildungsberichterstattung, 2012), macht deutlich, wie wichtig und notwendig eine gut strukturierte, systematische Sprachvermittlung ist. Hier sollte der Erwerb deutscher grammatikalischer Strukturen und Paradigmen immer schwerpunktmäßig behandelt werden, wobei der Genuserwerb sicherlich einen Kernbereich darstellt.

Das deutsche Genus-Kasus-System ist gekennzeichnet durch seine hohe Komplexität, welche zum einen aufgrund der Formenüberschneidung gegeben ist. So sind bestimmte Markierungen über Genusklassen hinweg identisch, wie beispielsweise die Verwendung des Artikels „der“ sowohl für maskuline Nomen im Nominativ, als auch für feminine Nomen im Genitiv und Dativ. Zum anderen macht die geringe Verfügbarkeit von sogenannten „noun cues“, die deutsche Genuszuweisung schwerer erwerbbar als in Sprachen, die das Genus phonologisch transparenter markieren. Der Begriff „cues“ bezieht sich im Folgenden auf Hinweise, die den Genuserwerb positiv beeinflussen. Noun cues befinden sich am Nomen selbst (z.B. bestimmte Endungen des Nomens) und erleichtern den Genuserwerb. In vielen Sprachen kann mit Hilfe der noun cues

eine Genuskategorie für einen Großteil der Nomen eindeutig zugewiesen werden. In einigen Sprachen, wie beispielsweise dem Spanischen, Italienischen oder Russischen treten die noun cues als bestimmte Endungen am Nomen (z.B. -a, -o) auf. Diese Endungen weisen klar auf ein Genus hin und erleichtern somit den Genuserwerb. Empirische Arbeiten konnten zeigen, dass diese noun cues den Erwerb von artifiziellen oder natürlichen Genuskategorien erleichtern (z.B. Brooks et al., 1993; Frigo & MacDonald, 1998; Kempe & Brooks, 2008).

Für das Deutsche ergibt sich diesbezüglich allerdings das Problem, dass deutsche Nomen nur in geringem Maße über noun cues verfügen. Obwohl eine Reihe von phonologischen und semantischen Regeln für die Genuszuweisung deutscher Nomen postuliert wurde (z.B. Köpcke, 1982; Wegener, 1995), gibt es für jede dieser Regeln eine Vielzahl von Ausnahmen, welche das System undurchsichtig erscheinen lassen. Beispielsweise besagt eine von Wegener (1995) aufgestellte Genuszuweisungsregel für das Deutsche, dass Mehrsilbern, die auf -el enden, meist das Maskulinum zugeordnet wird. Bei Vergleichen wie „Löffel“ (M) vs. „Gabel“ (F), „Klingel“ (F), „Kabel“ (N) etc., fällt auf, dass diese Regel für viele (hochfrequente) Nomina des Deutschen nicht gilt. Wenn man die deutschen Genuszuweisungsregeln (Wegener, 1995) auf den frühen Worterwerb überträgt, verringern sich die Fälle, in denen regelbasierende Genuszuweisungen möglich sind, noch einmal. Festzuhalten ist also, dass ein Lerner des Deutschen sich nur unzureichend an phonologischen Genuszuweisungsregeln orientieren kann.

Eine weiterführende Frage ist demnach, ob es weitere „cues“ oder Hinweise gibt, die den Genuserweb bei Sprachen ohne offensichtliche noun cues günstig beeinflussen. Taraban (2004) konnte in einer Experimentalreihe zeigen, dass die morphosyntaktische Umgebung des Nomens genau diese Aufgabe übernehmen kann, wenn man sie in einer lernerfreundlichen Weise präsentiert. Als sogenannte „syntactic cues“ werden Formen bezeichnet, die auf ein Genus hinweisen, z.B. die bestimmten Artikel eines maskulinen Nomens im Singular (der, des, dem, den) und seine zugehörigen maskulinen Pronomen im Singular (er, sein, ihm, ihn). Wenn diese Formen systematisch kombiniert werden, eröffnet sich für den Lerner die Möglichkeit, zu erfahren, welche Formen zusammengehören und auf ein Genus referieren. Taraban (2004, Experimente 3-5) arbeitete in seiner Studie mit einer artifiziellen Sprache, in der die artifiziellen Nomen über keinerlei phonologische noun cues verfügten. Er konnte zeigen,

dass der Erwerb genusartiger Kategorien durch eine solche systematische Kombination kritischer Formen erleichtert werden kann. Tarabans (2004) Forschung zeigt somit, dass syntactic cues durch eine gruppierte, geblockte Darbietung optimiert werden können, was wiederum den Lernerfolg positiv beeinflusst.

Im natürlichen Sprachgebrauch findet sich eine solche hochstrukturierte, gruppierte Darbietung von syntactic cues in den verschiedensten Formen von Sprachspielen. In vielen Sprachspielen, wie beispielsweise Liedern, Reimen oder Gedichten, rücken sprachliche Schwerpunkte bzw. die Kombination dieser unbewusst in den Vordergrund. So verknüpfen Sprachspiele z.B. morphosyntaktische Formen, die alle auf ein Nomen referieren und dadurch in ihrer Kombination auf die Genuskategorie Aufschluss geben können (vgl. z.B. 3a) bis 3c) in Abschnitt 4.1). Ein strukturierter, geblockter Input sprachlicher Formen und Strukturen kann als typisches Kennzeichen von Sprachspielen gelten. Weitere zentrale Charakteristika sind Melodie und Reim, die in vielen Sprachspielen sogar kombiniert auftreten.

Der positive Effekt von Musik für den Spracherwerb wurde durch eine Vielzahl an Studien nachgewiesen. So konnte unter anderem nachgewiesen werden, dass musikalische mit sprachlichen Fähigkeiten einhergehen (z.B. Anvari et al., 2002) und dass musikalisches Training die Entwicklung wichtiger mündlicher sowie schriftsprachlicher Fähigkeiten unterstützen kann (z.B. Ho et al., 2003; Moreno et al., 2009). Darüber hinaus zeigen Studien, dass sprachliches Lernen erleichtert werden kann, wenn dieses mit Melodie verknüpft wird (z.B. Schön et al., 2008). Die Verbindung zwischen Reimfähigkeiten und Sprachfähigkeiten konnte für verschiedene zentrale Bereiche des Spracherwerbs nachgewiesen werden (z.B. Foy & Mann, 2001; Wood & Terrell, 1998a), so dass eine positive Verbindung zwischen Musik- und Reimfähigkeiten auf der einen Seite und Spracherwerbsprozessen auf der anderen Seite angenommen wird.

Sprachspiele kombinieren einen hochstrukturierten sprachlichen Input mit Reim und Melodie und können als optimale Basis für die Förderung sprachlicher Fähigkeiten gelten (z.B. E. Belke & Belke, 2006). Ausgehend von Sprachspielen als einer Möglichkeit, Sprache systematisch zu fördern, ist es Ziel der vorliegenden Arbeit, wichtige Darbietungsformen von Sprachspielen psycholinguistisch zu untersuchen. Dabei stehen die oben vorgestellten

Darbietungsformen (Inputstrukturierung, Reim, Melodie) im Vordergrund, denn diese gelten als zentrale Merkmale von Sprachspielen.

Zum einen widmet sich die vorliegende Arbeit der Frage, ob die systematisch strukturierte (geblockte) Darbietung morphosyntaktischer Strukturen den Erwerb genusartiger Klassen erleichtert. Dafür wurden Probanden mit einer hochstrukturierten Inputdarbietung (Experiment 1 sowie Teile von Experiment 3) oder einer randomisierten Inputdarbietung (Experiment 2 sowie Teile von Experiment 3) trainiert. Die Probanden erwarben ein artifizielles genus-kasusartiges System, welches über keinerlei noun cues verfügte. Zum anderen überprüft die vorliegende Trainingsstudie den Einfluss der Variablen Reim und Melodie auf den Erwerb der strukturierten und randomisierten Inputdarbietung. Hierzu wurden die beiden Inputmethoden mit Reim- und/oder Melodiesstrukturen kombiniert.

Die zentrale Fragestellung der vorliegenden Arbeit ist, wie sich Darbietungsformen, die typischerweise im Sprachspiel vorliegen, auf den Erwerb genusartiger Klassen auswirken. Sprachspiele werden vorwiegend zur Sprachförderung mit Kindergarten- und Schulkindern eingesetzt. Da die Sprachförderung idealerweise so früh wie möglich beginnt und vor allem in jungen Jahren implizite Lernprozesse greifen, arbeitet die vorliegende Studie mit einer impliziten Vermittlungsituation, in der auf explizite Regelerklärungen verzichtet wird.

In Kapitel 2 dieser Arbeit diskutiere ich Studien, die sich mit der Rolle von semantischen und phonologischen noun cues beschäftigen. Wie bereits erwähnt zeigt die Forschung, dass noun cues den Erwerb genus(artiger) Kategorien erleichtern. Dies wurde für den Erwerb von artifiziellen sowie natürlichen sprachlicher genus(artigen) Kategorien gezeigt. Neben der Wirkungsweise der noun cues erläutere ich in diesem Zusammenhang die Besonderheit des Deutschen, welches nur über eine geringe Anzahl von noun cues verfügt. Diesbezüglich wird Literatur vorgestellt, die die umfassende Anwendbarkeit von noun cues im Spracherwerb des Deutschen in Frage stellt.

Basierend auf der Tatsache, dass noun cues für den Erwerb des deutschen Genussystems als nicht ausreichend angenommen werden können, konzentriert sich der zweite Teil von Kapitel 2 auf die Möglichkeit, den morphosyntaktischen Regelerwerb durch syntactic cues zu unterstützen. Wie bereits erwähnt konnte Taraban (2004) in seiner Studie zeigen, dass syntactic

cues dahingehend optimiert werden können, dass sie systematisch gruppiert dargeboten werden. Diese geblockte Inputdarbietung ermöglichte es den Lernern, genusartige Klassen trotz fehlenden noun cues zu erwerben.

Kapitel 3 beschäftigt sich mit der Rolle von Musik und Reim für die Verarbeitung und den Erwerb von Sprache. Der erste Teil dieses Kapitels stellt die Ähnlichkeiten von Sprach- und Musikverarbeitung in den Mittelpunkt. Zahlreiche Studien belegen eine enge Verbindung von Sprache und Musik in den verschiedensten Bereichen (z.B. Saffran et al., 1996b, Saffran et al., 1999; Steinbeis & Koelsch, 2008). Die Verknüpfung der beiden Domänen konnte sowohl anhand von Leistungsmessungen sowie durch elektrophysiologische Daten belegt werden. Kinder scheinen Musik und Sprache zu einem stärkeren Grad parallel zu verarbeiten als Erwachsene (Koelsch et al., 2003). Darüber hinaus konnte gezeigt werden, dass Sprache und Musik nicht nur ähnlich verarbeitet werden, sondern sich in ihrer Verarbeitung auch gegenseitig beeinflussen können (z.B. Schön et al., 2010).

Der zweite Teil des Kapitels beschäftigt sich mit der Frage, in wie weit Musik Sprachentwicklung beeinflussen kann. Zum einen werden hier Befunde vorgestellt, die eine Verbindung zwischen musikalischen und sprachlichen Fähigkeiten belegen (z.B. Sleve & Miyake, 2006). Zum anderen erläutere ich Studien, die musikalischem Training eine wichtige Rolle bei Spracherwerbsprozessen zuweisen (z.B. Marin, 2009). Schließlich diskutiere ich, wie Musik direkt in den Spracherwerbsprozess eingreifen und diesen positiv beeinflussen kann. So wurde beispielsweise gezeigt, dass sprachliches Sequenzlernen durch die zusätzliche Darbietung von Melodie unterstützt werden kann (z.B. Schön et al., 2008, Thiessen & Saffran, 2009).

Im dritten Teil des Kapitels wird die Bedeutung von Reim für den Spracherwerb erläutert. Studien zeigen, dass Reimwahrnehmung in direktem Bezug zu verschiedenen zentralen Aspekten des Spracherwerbs steht und Sprachfähigkeiten zuverlässig voraussagen kann (z.B. Foy & Mann, 2001).

Nachdem in den vorstehenden Kapiteln die drei Darbietungsformen Inputstrukturierung, Melodie und Reim erörtert wurden, stellt Kapitel 4 das Sprachspiel in den Mittelpunkt; eine Sprachform, die typischerweise alle drei Darbietungsformen vereint. Einer einleitenden Definition des Begriffs „Sprachspiel“ folgt eine spezifische Beschreibung zentraler Merkmale von Sprachspielen sowie deren Auswirkungen für Sprachlernprozesse. Im zweiten

Teil des Kapitels diskutiere ich Vorteile für den Einsatz von Sprachspielen in der gezielte Sprachförderung und -vermittlung. Schließlich wird das Konzept der Generativen Textproduktion (vgl. z.B. G. Belke, 2012) vorgestellt. Dieses didaktische Konzept setzt Sprachspiele gezielt ein, um sprachliche Strukturen, Regeln und Paradigmen spielerisch erwerbbar zu machen und zu festigen. Die Grundsätze, Vorgehensweise und die Relevanz der Generativen Textproduktion werden hier erläutert.

Auf der Grundlage der vorangegangenen Kapitel präsentiere ich in Kapitel 5 die empirische Untersuchung. Wie bereits erwähnt überprüft die vorliegende Arbeit die Wirkungsweisen der drei Darbietungsformen Inputstrukturierung, Reim und Melodie beim Erwerb artifizieller genusartiger Subklassen. Dafür wurden drei Experimente mit deutschen und englischen Muttersprachlern durchgeführt. Im ersten Teil des Kapitels gehe ich auf Hintergründe, Motivation und Fragestellungen der Studie ein. Die beiden zentralen Fragestellungen sind dabei, welchen Einfluss a) die optimierte Darbietung von syntactic cues durch Inputstrukturierung und b) die gleichzeitige Darbietung von Reim- und/oder Melodiestrukturen auf den Erwerb artifizieller genusartiger Klassen hat. Daran schließt sich im zweiten Teil ein Überblick über die Studie an, in dem das artifizielle genus-kasusartige System beschrieben wird. In den folgenden drei Unterteilen des Kapitels erfolgt die Beschreibung, Ergebnisdarstellung und Diskussion der Experimente sowie deren vergleichende Ergebnisdarstellung und -interpretation.

In Experiment 1 wurden deutsche Muttersprachler mit einer hochstrukturierten, geblockten Inputdarbietung trainiert. Hierbei wurden alle Sätze, die auf dasselbe Nomen referieren, direkt nacheinander dargeboten. Auf diesem Wege konnten die Probanden alle genus-kasus-markierten Formen zu jeweils einem Nomen gemeinsam verarbeiten. Um den Einfluss der Variablen Reim und Melodie auf eine systematische Inputstrukturierung zu überprüfen, wurden die Probanden in einem von vier Modi trainiert, die sich auf die Präsentation der Sätze bezogen. Dabei wurden die Sätze entweder in Reimform (Modus Reim), Melodieform (Modus Melodie), Reim- und Melodieform (Modus Reim-und-Melodie) oder ganz ohne Reim- und Melodieformen (Modus Prosa) präsentiert. Die Ergebnisse zeigen, dass bei fehlenden noun cues eine hochstrukturierte Inputdarbietung ausreichend ist, um genusartige Kategorien zu erwerben. Die Probanden aller vier Modi zeigten gute bis sehr gute Ergebnisse

bezüglich Anwendung des erworbenen Materials sowie der Generalisierung des Paradigmas. Zudem war ein Vorteil des Modus Reim-und-Melodie ersichtlich. Probanden dieser Gruppe erzielten die besten Ergebnisse und hoben sich vor allem im Vergleich zu der Prosa-Gruppe deutlich ab. Eine systematische Inputstrukturierung kann folglich durch die Kombination von Reim und Melodie zusätzlich unterstützt werden und für einen größeren Lernerfolg sorgen. Dieses Ergebnis weist auf den positiven Effekt der Kombination aller drei Darbietungsformen hin, welche im natürlichen Sprachgebrauch beispielsweise im Sprachspiel zu finden ist.

In Experiment 2 wurden deutsche Muttersprachler mit einer unstrukturierten, randomisierten Inputdarbietung trainiert. Analog zu Experiment 1 wurde auch hier der Einfluss von Reim und Melodie durch Einteilung der Probanden in vier Gruppen getestet. Die Ergebnisse zeigen ein ähnliches, wenn auch weniger konstantes Bild im Vergleich zu dem aus Experiment 1. Die Kombination von Reim und Melodie führte auch hier zu den besten Ergebnissen. Der Einfluss von Reim und Melodie auf die randomisierten Inputdarbietung war allgemein schwächer als auf die strukturierte Inputdarbietung. Das Ergebnismuster war jedoch in beiden Inputmethoden vergleichbar. Bei einem Vergleich der Ergebnisse aus Experiment 1 und 2 wird ersichtlich, dass die systematische Inputstrukturierung, wie in Experiment 1 gegeben, den größten positiven Effekt auf den Erwerb genusartiger Subklassen hat. Probanden, die mit strukturiertem Input trainiert wurden, erzielten bessere Ergebnisse als Probanden, denen die Sätze in randomisierter Reihenfolge dargeboten wurden. Die Kombination von Reim und Melodie erwies sich auch in der übergreifenden Analyse als unterstützend, vor allem in Zusammenhang mit einer strukturierten, geblockten Inputdarbietung: Probanden dieser Gruppe (geblockt, Reim-und-Melodie, siehe Experiment 1) konnten in allen Tests die besten Ergebnisse erzielen.

Experimente 1 und 2 arbeiteten mit deutschen Muttersprachlern, die bereits während ihres Erstspracherwerbs ein komplexes Genus-Kasus-System erworben haben. Es ist nicht auszuschließen, dass dieses Wissen über die Systematik und den Aufbau eines hochkomplexen morphosyntaktischen Systems ihnen (unbewusst) den Erwerb eines ähnlich systematischen Systems erleichtert hat. Um herauszufinden, inwieweit sprachliches Vorwissen den Erwerb des artifiziellen Systems begünstigen kann und ob die überprüften

Darbietungsformen sprachunabhängig greifen können, arbeitete Experiment 3 mit englischen Muttersprachlern. Um zu überprüfen, ob die Darbietungsformen einen ähnlichen Einfluss auf den Lernerfolg bei deutschen und englischen Muttersprachlern haben, arbeitete Experiment 3, analog zu Experiment 1, mit allen vier Gruppen (Prosa, Reim, Melodie, Reim-und-Melodie) der strukturierten/geblockten Methode. Zusätzlich wurde in Experiment 3 eine Gruppe aus der randomisierten Methode (vgl. Experiment 2) eingebunden. Aus den Ergebnissen der Experimente mit deutschen Muttersprachlern (Experiment 1 und 2) geht hervor, dass die Probanden der unstrukturierten Prosa-Gruppe die schlechtesten Ergebnisse erzielten. Diese Gruppe wurde daher als eine Art Baseline-Gruppe in die Arbeit mit englischen Muttersprachlern einbezogen. Experiment 3 umfasste somit insgesamt fünf Gruppen: alle vier Gruppen der strukturierten Methode sowie die Prosa-Gruppe der randomisierten Methode.

Die Ergebnisse aus Experiment 3 zeigen ein ähnliches Muster wie das der deutschen Muttersprachler. Auch hier zeigte sich die Inputstrukturierung als einflussreichste Variable. Dieses Ergebnis bestärkt die Annahme, dass strukturierter Input den Erwerb genusartiger Kategorien maßgeblich erleichtert. Dieser Effekt scheint sprachunabhängig zu greifen, auch dann, wenn kein vergleichbares System während des Erstspracherwerbs erworben wurde. Desweiteren erwies sich, parallel zu den deutschen Daten, dass die Kombination von Reim und Melodie als weitere unterstützende Komponente fungieren kann. Allerdings zeigte sich dieser Effekt stärker und konstanter in den deutschen als in den englischen Daten. Obwohl ein ähnliches Muster in beiden Sprachgruppen beobachtet werden konnte, weist dieses Ergebnis folglich auf Unterschiede zwischen den Ergebnissen deutscher und englischer Muttersprachler hin.

Eine sprachenvergleichende Analyse ergab, dass beide Sprachgruppen in ähnlichem Maße in der Lage waren, das erworbene Paradigma anzuwenden und auf unbekanntes Material zu übertragen. Allerdings zeigten die deutschen Probanden bessere Leistungen hinsichtlich der Aufgaben, die neben der Regelanwendung die Zuweisung der genusartigen Subklassen zu trainiertem Material erforderten. Ein weiterer Unterschied betrifft den Erwerb zusätzlicher Zweit- und Fremdsprachen. Wie anhand der Fragebogendaten ersichtlich, erwarben die deutschen Probanden mehr Fremdsprachen, genusmarkierende Sprachen inbegriffen, als die englischen Muttersprachler. Die Ergebnisse aus beiden Sprachgruppen weisen darauf hin, dass Spracherfahrung den Erwerb des

artifiziellen Paradigmas positiv beeinflusst. Ein weiteres, zentrales Ergebnis betrifft den Grad des expliziten Regelwissens. In beiden Sprachgruppen waren bessere Ergebnisse, wie sie in der strukturierten Methode zu finden waren, mit umfangreicherem expliziten Regelwissen über das artifizielle System verbunden. Strukturierter Input scheint also zu Regelwissen zu führen, welches, zumindest teilweise, ins Bewusstsein übergehen kann.

Kapitel 6 und 7 bieten eine Zusammenfassung der zentralen Ergebnisse sowie einen Ausblick hinsichtlich der Darbietungsformen, die für die Förderung morphosyntaktischer Fähigkeiten genutzt werden können. Desweiteren betrachte ich die vorliegende Studie kritisch und diskutiere die wichtigsten Schwachpunkte.

Zusammenfassend ist festzuhalten, dass in beiden Sprachgruppen strukturierter Input zu wesentlichem Lernzuwachs führte. Diese Form der Inputdarbietung beeinflusste den Erwerb genusartiger Kategorien am stärksten. Zudem kann die Kombination von Reim und Melodie den Lerneffekt unterstützen, vor allem, wenn sprachliches Lernen mit Hilfe von systematisch strukturiertem Input geschieht. Die drei Darbietungsformen Inputstrukturierung, Reim und Melodie scheinen also in ihrer Kombination besonders hilfreich. In alltäglichen Sprachfördersituationen findet sich diese Kombination unter anderem in vielen Formen des Spachspiels. Die Anwendung von Sprachspielen kann demnach aus psycholinguistischer Sicht für die Förderung von Spracherwerbsprozessen befürwortet werden. Die Befunde der vorliegenden Arbeit werden in Hinblick auf mögliche Konsequenzen für die Sprachförderung, vor allem von hochkomplexen morphosyntaktischen Systemen, abschließend diskutiert.

9 References

Andresen, H. (2007). Rollen, Regeln, Rätsel [Roles, Rules, Riddles]. In H. Andresen & F. Januschek (Eds.), *SpracheSpielen [LanguagePlay]* (pp. 27-44). Freiburg: Fillibach.

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83*, 111-130.

Arslan, F. (2005). Sprachvermittlung von Anfang an. Ein integratives Konzept zur Einführung der Artikel und ihrer Flexionen [Language promotion from the beginning. An integrative concept for the introduction of determiners and their inflections]. *Praxis Grundschule, 28*(2), 12-15.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Autorengruppe Bildungsberichterstattung [Authoring Group Educational Reporting] (2012). *Bildung in Deutschland 2012: Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf* [Education in Germany 2012 - An indicator-based report including an analysis of cultural education across the life span]. Bielefeld, DE: Bertelsmann. Retrieved from: http://www.bildungsbericht.de/daten2012/bb_2012.pdf (12.09.2013)

Belke, E., & Belke, G. (2006). Das Sprachspiel als Grundlage institutioneller Sprachvermittlung [Language play as the foundation of institutional language instruction]. In T. Becker & C. Peschel (Eds.), *Gesteuerter und ungesteuerter Grammatikerwerb [Directed and undirected grammar acquisition]* (pp. 174-200). Baltmannsweiler: Schneider Verlag Hohengehren.

Belke, G. (2007a). Verschiedene Sprachen - gemeinsames Spiel: Die Didaktik des Eigenen und des Fremden im multikulturellen Deutschunterricht [Various languages - same game: The didactic of personal and foreign aspects in a multicultural teaching of German]. In H. Andresen & F. Januschek (Eds.), *SpracheSpielen [LanguagePlay]* (pp. 90-118). Freiburg: Fillibach.

Belke, G. (2007b). *Poesie und Grammatik [Poetry and grammar].* Baltmannsweiler: Schneider Hohengehren.

Belke, G. (2008). *Mehrsprachigkeit im Deutschunterricht [Multilingualism in teaching of German].* Baltmannsweiler: Schneider Verlag Hohengehren.

Belke, G. (2012). *Mehr Sprache(n) für alle [More language(s) for everybody].* Baltmannsweiler: Schneider Verlag Hohengehren.

Bell, N. (2012). Formulaic language, creativity, and language play in a second language. *Annual Review of Applied Linguistics, 32*, 189-205.

Bordag, D., Opitz, A., & Pechmann, T. (2006). Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1090-1101.

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory and Cognition, 41*(2), 281-296.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76-95.

Brooks, P. J., Kempe, V., & Donachie, A. (2011). Second language learning benefits from similarity in word endings: Evidence from Russian. *Language Learning, 61*(4), 1142-1172.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics, 27*, 185-209.

Brown, S., & Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain and Cognition, 63*, 59-69.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research, 20*, 363-375.

Bryant, P., MacLean, M., & Bradley, L. (1990). Rhyme, language, and children's reading. *Applied Psycholinguistics, 11*(3), 237-252.

Chan, A. S., Ho, Y., & Cheung, M. (1998). Music training improves verbal memory. *Nature, 396*, 128.

Cook, G. (1997). Language play, language learning. *ELT Journal, 51*(3), 224-231.

Cook, G. (2000). *Language play, language learning*. Oxford: Oxford University Press.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex, 45*, 119-130.

Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research, 50*, 647-666.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics, 21*, 103-108.

Dabrowska, E., & Szczerbinski, M. (2006). Polish children's productivity with case marking: The role of regularity, type frequency, and phonological diversity. *Journal of Child Language, 33*(3), 559-597.

De Jong, N. (2005). Can second language grammar be learned through listening? *Studies in Second Language Acquisition, 27*, 205-234.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). WordGen: A tool for word selection and non-word generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments, & Computers, 36*, 488-499.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory and Cognition, 37*(1), 1-9.

Fonseca Mora, C. (2000). Foreign language acquisition and melody singing. *English Language Teaching Journal, 54*(2), 146-152.

Forkstam, C., & Petersson, C. M. (2005). Towards an explicit account of implicit learning. *Current Opinion in Neurology, 18*, 435-441.

Foy, J. G., & Mann, V. A. (2001). Does strength of phonological representations predict phonological awareness in preschool children? *Applied Psycholinguistics, 22*, 301-325.

Foy, J. G., & Mann, V. A. (2003). Home literacy environment and phonological awareness: Differential effects for phoneme awareness and rhyme awareness. *Applied Psycholinguistics, 24*(1), 59-88.

Foy, J. G., & Mann, V. A. (2006). Changes in letter sound knowledge are associated with development of phonological awareness in preschool children. *Journal of Research in Reading, 29*(2), 143-161.

Francois, C., & Schön, D. (2010). Learning of musical and linguistic structures: Comparing even-related potentials and behavior. *NeuroReport, 21*(14), 928-932.

Frieg, H., Hilbert, C., Belke, E., & Belke, G. (2012). Sprachförderung in ein- und mehrsprachigen Gruppen: Die generative Textproduktion [Language promotion in monolingual and multilingual groups: The generative Texproduktion]. *Sprachheilarbeit, 57*(3), 155-161.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*, 218-245.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS ONE, 5*(3), 1-12.

Griffin, E. A., & Morrison, F. J. (1997). The unique contribution of home literacy environment to differences in early literacy skills. *Early Child Development and Care, 127*(1), 233-243.

Grimm, H., Aktas, M., & Frevert, S. (2001). *SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder* [SETK 3-5. Language development test for three to five year old children; German version]. Göttingen: Hogrefe.

Gromko, J. E. (2005). The effect of music instruction on phonemic awareness in beginning readers. *Journal of Research in Music Education, 53*(3), 199-209.

Hallam, S. (2010). The power of music: Its impact on the intellectual, social and personal development of children and young people. *International Journal of Music Education, 28*(3), 269-289.

Haueis, E. (1985). Sprachspiele und die didaktische Modellierung von Wissensstrukturen [Language plays and the didactic implementation of knowledge structures]. In G. Stötzel (Ed.), *Germanistik - Forschungsstand und Perspektiven. Vorträge des deutschen Germanistentages [German Studies - Current research and perspectives]* (pp. 661-672). Berlin: de Gruyter.

Haueis, E. (2007). Nicht bloß zum Spaß: Spielen im Sprachunterricht [Not only for fun: Playing in language teaching]. In H. Andresen & F. Januschek (Eds.), *SpracheSpielen [LanguagePlay]* (pp. 119-135). Freiburg: Fillibach.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 269-286.

Ho, Y., Cheung, M., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology, 17*(3), 439-450.

Hoffmann, R., & Weis, I. (2011). *Deutsch als Zweitsprache - alle Kinder lernen Deutsch [German as a second language - all children learn German].* Berlin: Cornelsen.

Hoien, T., Lundberg, I., Stanovich, K. E., & Bjaalid, I. (1995). Components of phonological awareness. *Reading and Writing, 7*(2), 171-188.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*(2), 151-195.

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairmend also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20*(11), 1940-1951.

Jeuk, S. (2007). Schriftaneignung und Grammatisierung bei mehrsprachigen Kindern im Anfangsunterricht [Acquisition of written language and grammar learning in multilingual children in elementary instruction]. *Osnabrücker Beiträge zur Sprachtheorie, 73*, 35-55.

Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 627-640.

Kaltenbacher, R., & Klages, H. (2006). Sprachprofil und Sprachförderung bei Vorschulkindern mit Migrationshintergrund [Language profile and language promotion in preschool children with a migration background]. In B. Ahrenholz (Ed.), *Kinder mit Migrationshintergrund. Spracherwerb und Fördermöglichkeiten [Children with a migration background. Language acquisition and promotion opportunities]* (pp. 80-97). Freiburg: Fillibach.

Kauffeldt, L. (2011). *Musikalische Gestaltung im Projekt: Bilinguale Sprachverarbeitung unter Submersionsbedingungen. Eine psycholinguistische Aufarbeitung [Musical arrangement in the project: Bilingual language processing with conditions submersion. A psycholinguistic process]* (Bachelor thesis). Institut für Musikwissenschaft und Musikpädagogik, Westfälische Wilhelms-Universität Münster, Münster.

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning, 58*(4), 703-746.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*(1), 171-184.

Kilgour, A. R., Jakobson, L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediator of text and song recall. *Memory & Cognition, 28*, 700-710.

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology, 15*, 207-212.

Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schröger, E., & Friederici, A. D. (2003). Children processing music: Electric brain responses reveal musical competence and gender differences. *Journal of Cognitive Neuroscience, 15*(5), 683-693.

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., & Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition, 112*, 1-20.

Köpcke, K. M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache [Investigation of the gender system of contemporary German].* Tübingen: Niemeyer.

Köpcke, K. M., & Zubin, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen [Six principles for German gender assignment]. *Linguistische Berichte, 93*, 26-50.

Kramarz, V. (2007). *Die Pop Formeln [The pop formulars].* Bonn: Voggenreiter.

Lang, B. (2007). ‚la-la-la' und ‚ackabacka schu': Spielen mit Lauten vom Säuglingsalter bis ins Schulalter - eine entwicklungsorientierte Perspektive [‚la-la-la' und ‚ackabacka schu': Playing with sounds from infancy to school age - a developmental perspective]. In H. Andresen & F.

Januschek (Eds.), *SpracheSpielen [LanguagePlay]* (pp. 71-89). Freiburg: Fillibach.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28*, 255-277.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*(5), 540-545.

Mann, V. A., & Foy, J. G. (2003). Speech development, phonological awareness, and letter knowledge in preschool children. *Annals of Dyslexia, 53*, 149-173.

Mann, V. A., & Foy, J. G. (2007). Speech development patterns and phonological awareness in preschool children. *Annals of Dyslexia, 57*, 51-74.

Maratsos, M. (1983). Some current issues in the study of the acquisition of grammar. In P. Mussen (Ed.), *Handbook of child psychology* (Vol. III, pp. 707-786). New York: Wiley.

Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. E. Nelson (Ed.), *Children's language* (Vol. 2, pp. 127-214). New York: Gardner Press.

Marin, M. M. (2009). Effects of early musical training on musical and linguistic syntactic abilities. *The Neurosciences and Music III - Disorders and Plasticity: Annals of the New York Academy of Sciences, 1196,* 187-190.

Mathews, R.C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1083-1100.

Mc Mullen Jonaitis, E., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science, 33*, 951-968.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception, 21*(3), 289-311.

Moreno, S. (2009). Can music influence language and cognition? *Contemporary Music Review, 28*(3), 329-345.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebal Cortex, 19*(3), 712-723.

Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Osterhout, L., & Holcomb, P. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8*, 413-437.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*, 717-733.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*(2), 227-244.

Röhner, C. (2005). *Erziehungsziel Mehrsprachigkeit. Diagnose von Sprachentwicklung und Förderung von Deutsch als Zweitsprache [Educational goal monolingualism: Diagnostics of language development and promotion of German as a second language].* Weinheim: Juventa.

Rösch, H. (2003). „fom Afrika bies zu Berlin“ - Deutsch als Zweitsprache (DaZ) in der Grundschule [„From Africa to Berlin“ (original utterance includes grammatical and orthographic errors) - German as a Second Language in elementary school]. *Grundschule, 5/2003*, 44-46.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*(1), 1-29.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493-515.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Johnson, E., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Sanches, M., & Kirshenblatt-Gimblett, B. (1976). Children’s traditional speech play and child language. In B. Kirshenblatt-Gimblett (Ed.), *Speech play.*

Research and resources for studying linguistic creativity (pp. 65-110). Pennsylvania: University of Pennsylvania Press.

Schellenberg, E. G., & Trehub, S. (1999). Culture-general and culture-specific factors in the discrimination of melodies. *Journal of Experimental Child Psychology, 74*, 107-127.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106,* 975-983.

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage, 51*, 450-461.

Schröder, A., Kauschke, C., & De Bleser, R. (2003). Messungen des Erwerbsalters für konkrete Nomina [Measurements on the age of acquisition for concrete nouns]. *Neurolinguistik, 17*(2), 83-114.

Schwichtenberg, B., & Schiller, N. O. (2004). Semantic gender assignment regularities in German. *Brain and Language, 90*, 326-337.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115*(2), 163-196.

Seigneuric, A., Zagar, D., Meunier, F., & Spinelli, E. (2007). The relation between language and cognition in 3- to 9-year-olds: The acquisition of grammatical gender in French. *Journal of Experimental Child Psychology, 96*, 229-246.

Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445-460.

Shimpi, P. M., Gámez, P. B., Huttenlocher, J., & Vasilyeva, M. (2007). Syntactic priming in 3- and 4-year-old children: Evidence for abstract representations of transitive and dative forms. *Developmental Psychology, 43*(6), 1334-1346.

Siegmüller, J., & Kauschke, C. (2006). *Patholinguistische Therapie bei Sprachentwicklungsstörungen [Patholinguistic therapy with specific language impaired children].* München: Elsevier.

Sleve, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17*(8), 675-681.

Smith, L. B., Kemler Nelson, D., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition, 52*, 23-54.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174-215.

Stainthorp, R., & Hughes, D. (1998). Phonological sensitivity and reading: Evidence from precocious readers. *Journal of Research in Reading, 21*(1), 53-68.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*, 1169-1178.

Szagun, G. (2004). Learning by ear: On the acquisition of case and gender marking by German-speaking children with normal hearing and with cochlear implants. *Journal of Child Language, 31*(1), 1-30.

Szagun, G., Stumper, B., Sondag, N., & Franik, M. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language, 34*(3), 445-471.

Taraban, R. (2004). Drawing learners' attention to syntactic context aids gender-like category induction. *Journal of Memory and Language, 51*, 204-216.

Taraban, R., & Kempe, V. (1999). Gender processing in native and nonnative Russian speakers. *Applied Psycholinguistics, 20*, 119-148.

Thiessen, E. D., & Saffran, J. R. (2009). How the melody faciliates the message and vice versa in infant learning and memory. *The Neurosciences and Music III - Disorders and Plasticity: Annals of the New York Academy of Sciences, 1196*, 225-233.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*(1), 1-42.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology, 102*, 120-129.

Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading, 29*(3), 334-348.

Tillmann, B., Bharucha, J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107*(4), 885-913.

Trainor, L., & Trehub, S. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 394-402.

Vahle, F. (1988). *Das Anne Kaffeekanne Liederbuch. Lieder zum Singen, Spielen und Tanzen [The Anne Kaffeekanne songbook. Songs for singing, playing and dancing].* Hamburg: Nikol.

van den Bos, E., & Poletiek, F. H. (2010). Structural selection in implicit learning of artificial grammars. *Psychological Research, 74*, 138-151.

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: Insight from grammatical-SLI*. Trends in Cognitive Sciences, 9*(2), 53-59.

von Lehmden, F. M., Kauffeldt, J., Belke, E., & Rohlfing, K. J. (2013). Das Vorlesen von Kinderbüchern als implizites Mittel zur Sprachförderung im Bereich Grammatik [Reading to children promotes implicit grammar acquisition]. *Praxis Sprache, 58*(1), 18-27.

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1471-1485.

Wegener, H. (1995). *Die Nominalflexion des Deutschen - verstanden als Lerngegenstand [The nominal inflection of German - seen as a learning matter].* Tübingen: Niemeyer.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*, 269-304.

Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing, 15*, 179-206.

Wood, C. (2006). Metrical stress sensitivity in young children and its relationship to phonological awareness and reading. *Journal of Research in Reading, 29*(3), 270-287.

Wood, C., & Terrell, C. (1998a). Pre-school phonological awareness and subsequent literacy development. *Educational Psychology, 18*(3), 253-274.

Wood, C., & Terrell, C. (1998b). Poor readers' ability to detect speech rhythm and perceive rapid speech. *British Journal of Developmental Psychology, 16*, 397-413.

Zubin, D. A., & Köpcke, K. M. (1984). Affect classification in the German gender system. *Lingua, 63*, 41-96.

Appendices

Appendix A

Artificial Object Nouns Used in Experiments 1, 2, and 3

Item number	Object Noun	Number of Phonemes in German	Number of Phonemes in English	Summated Bigram Frequency (German)	Summated Bigram Frequency (English)
1	elom	4	4	7787	5707
2	teos	4	4	8214	5868
3	riun	4	4	10290	5354
4	unak	4	5	9259	4032
5	etef	4	4	10409	6985
6	ilil	4	4	7599	6921
7	deluf	5	5	11848	5971
8	eitak	4	4	16832	5848
9	besak	5	5	11047	6312
10	refam	5	5	10431	7096
11	gunal	5	5	12615	8101
12	jetem	5	6	10741	7626
13	gitun	5	6	13482	6372
14	lelop	5	5	11810	9436
15	urets	5	6	12582	8644
16	osret	5	5	11122	7712
17	ketiz	6	5	10727	8514
18	atunz	6	5	12051	7641
19	filan	5	5	1438	9668
20	molun	5	5	10379	5532
21	tekes	5	5	14104	9613
22	geop	4	4	6572	2849
23	tetak	5	5	13261	9152
24	inut	4	4	7084	8991

25	melam	5	5	12564	7841
26	witim	5	5	11101	9560
27	akest	5	5	13904	9547
28	rebif	5	5	10397	6458
29	unol	4	5	8837	4660
30	noref	5	5	11274	9604
31	astop	5	5	10505	8898
32	imres	5	5	11300	9785
33	abun	4	4	8898	4011
34	runot	5	5	10805	5029
35	elwan	5	5	11828	7585
36	nelun	5	5	16103	8570

Note. Items 31 to 36 (novel-new object nouns) were not used during the training procedure. Items 25 to 30 (novel-old object nouns) were only used during the vocabulary training, but not during the syntax training.

Appendix B

Questionnaire - German version

Bedingung: _____ Versuchsperson: _____ Version: _____

1. Alter: _____ 2. Geschlecht: m ☐ w ☐

3. Außer Deutsch spreche ich

Englisch	seit ___ Jahren und verwende es	häufig ☐	gelegentlich ☐	selten ☐
Französisch	seit ___ Jahren und verwende es	häufig ☐	gelegentlich ☐	selten ☐
Spanisch	seit ___ Jahren und verwende es	häufig ☐	gelegentlich ☐	selten ☐
Italienisch	seit ___ Jahren und verwende es	häufig ☐	gelegentlich ☐	selten ☐
_____	seit ___ Jahren und verwende es	häufig ☐	gelegentlich ☐	selten ☐
_____	seit ___ Jahren und verwende es	häufig ☐	gelegentlich ☐	selten ☐

4. Wie musikalisch schätzt du dich ein?

musikalisch ☐ eher musikalisch ☐ eher unmusikalisch ☐ unmusikalisch ☐

5. Hast Du während des Trainings und im Test bemerkt, dass die Sätze bestimmten Regeln folgten?

Ja ☐ Nein ☐ Falls ja, bitte versuche, die Regeln zu beschreiben:

_____

_____

_____

_____

_____

6. Wie leicht ist es dir in der heutigen Sitzung gefallen, eigenständige Sätze zu produzieren?

leicht ☐ eher leicht ☐ eher schwer ☐ schwer ☐

7. Fandest Du die Produktion eigenständiger Sätze In Sitzung 2 und 3 beim Sprechen einer ganzen Strophe einfacher, als wenn Du die Sätze einzeln („gemischt“) produzieren musstest?

Ja ☐ Nein ☐ Die Aufgaben waren gleich schwierig ☐

8. Wie hast du die Sätze in den Produktionsaufgaben im Training gebildet?

nach Gefühl ☐ mir war die Regel bewusst ☐

anders ☐, und zwar

_____

9. Wie hast du die Sätze in den Produktionsaufgaben im Test gebildet?

nach Gefühl ☐ mir war die Regel bewusst ☐

anders ☐, und zwar

_____

10. Wie hast Du die Auswahlaufgabe in Sitzung 2 gelöst, in der Du zu einer Äußerung die richtige Handlung auswählen solltest?

nach Gefühl ☐ mir war die Regel bewusst ☐

anders ☐, und zwar

_____

11. Wie hast Du die Satz- und Strophenbeurteilungsaufgabe im Test (heute) gelöst?

nach Gefühl ☐ mir war die Regel bewusst ☐

anders ☐, und zwar

_____

Questionnaire - English version

Condition: _____ Participant: _____ Version: _____

1. Age: _____ 2. Sex: m ☐ f ☐

3. Besides English, I can speak the following languages:

French	for ___ years and I use it	often ☐	sometimes ☐	seldom ☐
Spanish	for ___ years and I use it	often ☐	sometimes ☐	seldom ☐
Italian	for ___ years and I use it	often ☐	sometimes ☐	seldom ☐
_____	for ___ years and I use it	often ☐	sometimes ☐	seldom ☐
_____	for ___ years and I use it	often ☐	sometimes ☐	seldom ☐

4. From your point of view, how musical are you?

musical ☐ rather musical ☐ rather unmusical ☐ unmusical ☐

5. While participating in this study, did you notice that the sentences followed certain rules?

Yes ☐ No ☐ If yes, please try to describe the rules:

_____

_____

_____

_____

_____

6. In the session today, how easy was it for you to produce sentences on your own?

easy ☐ rather easy ☐ rather difficult ☐ difficult ☐

7. In session 2 and 3, was it easier for you to produce the sentences in forms of groups than to produce the sentences in random (“mixed”) order?

Yes ☐ No ☐ There was no difference. ☐

8. How did you produce the sentences in the training sessions (sessions 2 and 3)?

By instinct ☐ I was aware of the rules ☐

differently ☐, namely

_____

9. How did you produce the sentences in the test session (today)?

By instinct ☐ I was aware of the rules ☐

differently ☐, namely

_____

10. How did you solve the choice task in session 2, where you were asked to choose the correct sentence for an animated event?

By instinct ☐ I was aware of the rules ☐

differently ☐, namely

_____

11. How did you solve the choice tasks in the test session (today?)

By instinct ☐ I was aware of the rules ☐

differently ☐, namely

_____

Curriculum Vitae Johanna Bebout, geb. Kauffeldt

Akademische Laufbahn:

10/2009 – 07/2015	Ruhr-Universität Bochum, Promotionsprojekt im Fach Linguistik (Prof. Dr. Eva Belke, Prof. Dr. Ralf Klabunde), Titel der Dissertation: „Can language play promote morphosyntactic learning? A study about the impact of rhyme, melody and input structuring on gender-like category induction."
04/2008 – 09/2009	Universität Bielefeld, Studiengang Klinische Linguistik (M.Sc.), Titel der Masterarbeit: "Das Vorlesen von Kinderbüchern als Mittel zur Sprachförderung im Bereich Grammatik: Eine Evaluationsstudie zum Passiverwerb". Supervision durch Prof. Dr. Eva Belke (Ruhr-Universität, Bochum) & Dr. Katharina Rohlfing (Universität Bielefeld)
10/2004 – 03/2008	Universität Bielefeld, Studiengang Klinische Linguistik (B.Sc.), Titel der Bachelorarbeit: "Falldarstellung eines Patienten mit Broca-Aphasie und Sprechapraxie". Supervision durch Prof. Dr. Gert Rickheit (Universität Bielefeld) & Prof. Dr. Dr. Horst Müller (Universität Bielefeld)
08/2001 – 06/2004	Abitur; Friedrich-von-Bodelschwingh-Gymnasium Bethel, Bielefeld

Berufliche Laufbahn:

seit 06/2013	Sprachtherapeutin im Cochlear Implant Centrum Ruhr, Essen
10/2009 – 09/2013	Wissenschaftliche Mitarbeiterin im Forschungsprojekt „Bilinguale Sprachverarbeitung unter Submersionsbedingungen: Eine psycholinguistische Aufarbeitung." (Projektleitung: Prof. Dr. Eva Belke), Sprachwissenschaftliches Institut, Ruhr-Universität Bochum
02/2009 – 09/2009	Sprachtherapeutische Tätigkeit, Logopädische Praxis Gaisendrees, Gütersloh
04/2009 – 07/2009	Individuelle DaZ (Deutsch als Zweitsprache) – Sprachförderung mit dem Konzept der Generativen Textproduktion in der 5.Klasse, in Zusammenarbeit mit dem Amt für Integration und interkulturelle Angelegenheiten – RAA, Bielefeld und der Realschule Heepen, Bielefeld

06/2008 – 09/2008 Studentische Hilfskraft im Projekt “Bilinguale Sprachverarbeitung unter Submersionsbedingungen: Eine psycholinguistische Aufarbeitung. (Projektleitung: Dr. Eva Belke), Universität Bielefeld

Publikationen:

von Lehmden, F., Bebout, J. & Belke, E. (in Vorbereitung). Effizienter Einsatz von Inputoptimierung beim Vorlesen in der Sprachförderung. In E. Gressnich, C. Müller & L. Stark (Hrsg.), *Lernen durch Vorlesen.* Narr Dr. Gunter.

Frieg, H. & Bebout, J. (2013). Spielerisch sprechen und schreiben lernen. *Praxis Fördern, 6/2013,* 8-19.

von Lehmden, F., Kauffeldt, J., Belke, E., & Rohlfing, K. (2013). Das Vorlesen von Kinderbüchern als implizites Mittel zur Sprachförderung im Bereich Grammatik. *Praxis Sprache*, *58,* 18-27.

Frieg, H., Belke, E., Belke, G., Hoffmann, R., Bebout, J., Kauffeldt, L. & Kirschke, C. (2014). *Dschungeltanz & Monsterboogie. Lieder zur systematischen Sprachvermittlung im Vor- und Grundschulalter.* Baltmannsweiler: Schneider Hohengehren.

Kauffeldt, L., Kirschke, C., Bebout, J., Frieg, H., Belke, E., Hoffmann, R. & Belke, G. (2014). *Dschungeltanz & Monsterboogie. Singen und Spielen mit Sprache.* Baltmannsweiler: Schneider Hohengehren.

Frieg, H., Kauffeldt, J. & Belke, E. (2010). Einleitende Gedanken zur Bedeutung der Mehrsprachigkeit. In Türkischer Lehrerverein Nordrhein-Westfalen e.V. (Hrsg.), *Dokumentation des Vorlesewettbewerbs in Deutsch und Italienisch sowie in Deutsch und Türkisch im Schuljahr 2009/10*, 19-20.