Zeitschrift: Bulletin suisse de linguistique appliquée / VALS-ASLA

Herausgeber: Vereinigung für Angewandte Linguistik in der Schweiz = Association suisse de linguistique appliquée

Band: - (2022)

Heft: 115: Implizites und explizites sprachliches Wissen und Vermitteln : interdisziplinäre Fragestellungen - diszipinäre Zugänge = Implicit and explicit linguistic knowledge and instruction : interdisciplinary issues - disciplinary approaches

Artikel: Help or hindrance? : Effects of additional morphosyntactic cues on the acquisition of gender-like subclasses in an artificial language by 4- to 6-year-olds

Autor: Kuba, Sarah / Braun, Pia Marie / Belke, Eva

DOI: https://doi.org/10.5169/seals-1062982

Help or hindrance? Effects of additional morphosyntactic cues on the acquisition of gender-like subclasses in an artificial language by 4- to 6-year-olds

Sarah KUBA, Pia Marie BRAUN & Eva BELKE[1]
Sprachwissenschaftliches Institut
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum
sarah.kuba@rub.de; pia.m.braun@rub.de; eva.belke@rub.de

Der Erwerb des Genus im Deutschen ist oft langwierig. Zwar tragen viele Nomen semantische, phonologische oder morphologische Hinweise auf ihr Genus, aber diese sind teilweise widersprüchlich und haben viele Ausnahmen. Zuverlässiger sind Hinweise aus dem syntaktischen Kontext des Nomens. Zwar sind auch sie als isolierte Hinweise aufgrund der Formenüberschneidungen in den Paradigmen der Nominalflexion nicht eindeutig, aber kombinierte Hinweise aus mehreren Paradigmen geben eindeutige Hinweise auf das Genus eines Nomens. Allerdings stellen sie eine größere Belastung des phonologischen Kurzzeitgedächtnisses dar. In einer Studie im *Artificial Language Learning*-Paradigma untersuchten wir, wie gut Kinder zwischen 4;6 und 6;6 Jahren genusartige Subklassen in einer Kunstsprache erwerben, wenn der Input Hinweise auf die Subklassen aus nur einem artikelartigen künstlichen Markerparadigma oder zusätzlich Adjektivsuffixe enthielt. Nur mit beiden Markierungen waren die Zuordnungen der Markerkombinationen zu grammatischen Funktionen eindeutig. Die Leistungen der Kinder, deren Training Adjektivsuffixe enthielt, waren tendenziell schlechter als die der Vergleichsgruppe, aber qualitative Analysen ihrer Fehler legen nahe, dass nur sie verstanden hatten, dass es zwei verschiedene Subklassen gab. Wir nehmen an, dass gut wahrnehmbare Adjektivsuffixe ihre Aufmerksamkeit auf die relevanten Hinweiskombinationen im Markerparadigma lenkten.

Stichwörter:
Spracherwerb, Genuserwerb, implizites Lernen, Inputoptimierung.

Keywords:
language acquisition, gender acquisition, implicit learning, input optimization.

1. Introduction

Implicit learning is key to language acquisition (e.g. Ullman & Pierpont 2005; Savage et al. 2006). It features in the acquisition of regularities in sequences of stimuli as well as in the acquisition of linguistic categories and subclasses based on distributional information in the environment (Culbertson & Schuler 2019). However, some linguistic domains are hard to acquire implicitly due to the elusiveness of relevant cues. One such domain are gender subclasses of nouns in

[1] This research was funded by a grant from the Volkswagen Foundation awarded to Eva Belke. It is part of the collaborative research project "Literacy as the key to social participation: Psycholinguistic perspectives on orthography instruction and literacy acquisition (Litkey)" (grant no. I/89 479) funded in the research initiative "Key Issues for Research and Society". We thank Carola Sicken for her invaluable help in data acquisition and coding.

German. As we will review shortly, there are only few consistent semantic or form cues to the grammatical gender of most early acquired nouns. Syntactic cues, such as definite articles, are often ambiguous as well due to a high degree of syncretism in the nominal inflection system. The opacity resulting from this can only be resolved by combining multiple markers to yield unambiguous cues to grammatical gender. Given this, it would be desirable to facilitate children's acquisition of gender subclasses by directing their attention to unique marker combinations. One way of doing so is to optimize the presentation of the input. Studies with adult learners suggest that texts in children's songs and language play group present relevant grammatical markers in such a way that learners can use them in acquiring grammatical subclasses (Bebout & Belke 2017; see also Cook 2000). Optimizing the linguistic input in this way in order for learners to pick up on relevant syntactic cues is a means to making systematic use of learners' implicit learning resources, which is why we refer to it as *implicit teaching* (E. Belke & G. Belke 2006).

We hypothesize that texts in children's books, too, can be used to teach implicitly the assignment of nouns to gender-like subclasses: When novel actors or entities are introduced in a story, they are typically referred to by an indefinite NP. When they are mentioned again shortly after, they are referred to by a definite NP with the same head. If the NPs contain an adjective, its inflection depends, among other factors, on the definiteness of the NP. This way, various relevant syntactic markers (indefinite and definite articles and adjective inflections) are being combined to form unique cues to the grammatical gender of the head of the NP (von Lehmden et al., subm.). Note however, that this form of input optimization comes at the cost of combining multiple cues in one phrase. Extracting cues to grammatical gender from such complex phrases may be hampered by limitations of phonological short-term memory (STM) and phonological working memory (for a review, see Gathercole 2007).

In the present study, we used an artificial language learning (ALL) paradigm to investigate whether combining cues to grammatical gender subclasses helps or hinders their acquisition. Unlike previous studies, we worked with preschool children for several reasons: First, it is at this age that gender subclasses should ideally have been acquired at the latest (Ruberg 2015), so if they have not, an implicit teaching intervention would be a way to support children in the acquisition process. Second, German children typically start school and formal reading instruction at the age of 6, and we figured that to acquire orthography effectively and to become literate speakers/writers, it is essential that, prior to entering school, children acquire sufficient implicit knowledge of grammatical gender. It is key to anaphoric relations in written, literate texts and is hence of central relevance to literacy acquisition.

In the present study, we constructed two variants of an artificial language with two gender-like subclasses and trained one group of preschool children each with one of the variants. While the first variant included only one set of partially overlapping distributional cues to the two subclasses, the second variant featured combined, unambiguous cues to the two subclasses. We hypothesized that a) combining multiple cues might be more successful in drawing learners' attention to the existence of the gender-like subclasses and the morphosyntactic markers associated with them but that b) the increased phonological complexity associated with additional cues might hamper acquisition. To explore these hypotheses, we assessed whether the two groups of children differed in overall accuracy and in what way they differed in their use of grammatical markers, correct or otherwise.

In the remainder of this introduction, we will review when grammatical gender is typically acquired by learners of German as L1 or as L2 and what makes its acquisition so hard. After that, we will review ALL studies of the acquisition of gender-like subclasses and of the effects of input optimization on the acquisition.

1.1 Acquisition of Grammatical Gender in German

There are several longitudinal and cross-sectional studies on the acquisition of grammatical gender in children acquiring German as L1 or as L2, respectively. In reviewing the data from these studies, it is important to distinguish between analyses of utterances produced spontaneously *vs.* in tasks designed to elicit gender-marked forms in (pseudo-)nouns (e.g. Ruberg 2015).

A wide range of estimates has been reported for the age at which children with German as L1 achieve 90% accuracy in gender-marked nominal inflections. Based on a longitudinal corpus of utterances produced in free play sessions, Szagun et al. (2007) observed that children reached full acquisition by the age of 3. By contrast, Ruberg (2015), who tested 16 monolingual German children aged 3;0 to 5;3 in an experimental setting, observed full acquisition of indefinite articles for children aged 4;6 years on average, indicating that experimental settings are more sensitive in uncovering gaps in the children's knowledge of the grammatical gender of nouns. For definite articles, some children had not yet reached more than 90% accuracy by about 5;0 years and none of the age groups (3;0, 3;6, 4;0, 4;6, and 5;0 years) reached full acquisition of gender markings with attributive adjectives. Ulrich (2017) tested 968 children between the ages of 4 and 9 cross-sectionally in an elicitation task and found that only 27% of the 4-year-olds had reached the threshold of 90% accuracy. Most children were between six and seven when reaching 90% accuracy. Children acquiring German as L2 fully master gender assignment in German even later than that (Jeuk 2008; Ruberg 2015).

This review of the literature suggests that the acquisition of grammatical gender in German is hard for all learners (Ulrich 2017). This may be the case because many of the cues to grammatical gender are inconsistent. Various studies have established semantic, phonological, and morphological regularities for gender assignment in German nouns (e.g. Köpcke & Zubin 1984), also referred to as noun cues. However, as Szagun et al. (2007) have pointed out, young children do not have the ability to abstract common semantic features yet, so the semantic fields that Köpcke & Zubin (1984) established as relevant cues to gender assignment likely do not play a role in gender acquisition of children acquiring German as L1 (see Wegener 1995, for a parallel discussion of cue validity for L2 learners of German).

Szagun et al. (2007) found that, in their corpus of spontaneous utterances, children made more errors with words that were inconsistent with phonological regularities to gender assignment. However, Ulrich (2017) did not find any evidence that the children she tested used noun-internal principles of gender assignment, neither in the correct nor in their erroneous answers. For instance, *Banane* (banana.F), which ends in [ə], a strong cue for the feminine gender (Wegener 1995), was not processed more accurately than *Gans* (goose.F) or *Pferd* (horse.N), which include no overt cues to their gender, yielding 26.1%, 25.6% and 25.6% errors, respectively. *Gans* and *Pferd* are monosyllabic, which is often a cue for masculine gender (Wegener 1995), yet the children assigned the masculine gender erroneously less often than the neuter (*Gans*) and feminine (*Pferd*) gender.

Using 222 object nouns that had been rated for their age-of-acquisition by Schröder et al. (2003), Bebout & Belke (2017) examined the predictive values of the gender assignment rules put forward by Wegener (1995). Of the nouns with an age of acquisition of 3 years or younger (cf. Szagun et al. 2007), only 56% could be assigned the correct gender following Wegener's (1995) regularities. To summarize, it appears that in the early stages of the acquisition of grammatical gender, noun cues are of little help to children acquiring German as L1 or L2.

Unlike noun cues, syntactic cues, such as nominal inflectional morphemes accompanying a noun, can be reliable cues to grammatical gender. However, it is rarely possible to determine a noun's gender based on individual morphemes due to syncretism in the nominal inflection system. For instance, the indefinite article *ein* ('a') can co-occur with masculine and neuter nouns. Similarly, the definite article *der* ('the') can occur with masculine and feminine nouns in the nominative and dative case, respectively. Ambiguity inherent to syntactic cues may be reduced when several cue paradigms are combined, such as definite articles and adjective inflections. However, combined cues are longer and hence impose more load on phonological STM. It is well established that phonological short-term memory is an important predictor of vocabulary knowledge in 4- to 6-year olds (for review, see Gathercole 2007). Verhagen & Lesemann (2016) showed that in 5-year-old children who acquired Dutch as L1 or as L2, phonological short-term memory (as assessed by pseudoword repetition and word repetition tasks) was a significant predictor of their acquisition of both vocabulary and grammar in Dutch. Most children were not able to repeat nonwords with more than a few syllables (Verhagen & Lesemann, 2016, Table 1). Hence, adding one syllable to the training sequences of an ALL study can have substantial effects on the learning outcome.

1.2 Acquisition of Gender-Like Subclasses in Artificial Language Learning Studies

ALL paradigms have been used to study various aspects of language acquisition in children and adults (for review, see Culbertson & Schuler 2019). In the context of the acquisition of gender-like subclasses, the effects of individual types of cues and their properties on participants' acquisition of the subclasses can be studied. For instance, Brooks et al. (1993) conducted a study with adults and children (aged 9 to 10 years) who were exposed to one of two artificial languages. Both languages included two gender-like noun subclasses and six locative suffixes. The experimental language featured phonological similarities between most nouns of each gender-like subclass, while there were no such similarities in a control language. Brooks et al. found that adults and children trained with the experimental language acquired the gender-like subclasses and the associated marker paradigms, while participants of the control groups performed at chance. The authors took this to show that noun cues, such as the phonological cues used in their study, are indispensable for acquiring gender-like subclasses.

Taraban (2004) used the control language by Brooks et al. (1993) in experiments with adults to investigate whether the acquisition of gender-like subclasses for phonologically unmarked nouns was possible if the input was structured in a learner-friendly fashion. In one training condition, he employed a blocked input presentation such that all sentences pertaining to a given noun were presented in immediate succession. In a control condition, the sentences were presented in a random order during training. Participants trained with the blocked training regime had acquired the two gender-like subclasses and performed significantly better than the control group, who performed at chance, replicating the findings reported by Brooks and colleagues for this language. Taraban hypothesized that the blocked presentation drew learners' attention to the groups of morphemes that were relevant for the distinction between gender-like subclasses. This form of input optimization can be considered a form of implicit teaching.

Bebout & Belke (2017) extended Taraban's study (2004) by investigating whether the blocked input presentation would still be more effective than a random one if a) the system of gender-like subclasses included three rather than two gender-like subclasses and b) the paradigms of syntactic cues involved substantial marker overlap as seen in German. They replicated the finding that, at test, participants trained with a blocked presentation regime had acquired the gender-like subclasses while those trained with a random-order presentation regime performed at chance. In addition, the authors showed that compared to a presentation in prose, a rhymed, a melodic and a rhymed *and* melodic presentation enhanced the learning outcome substantially, especially in the groups who had been trained with a blocked training regime. The best results were obtainned by learners trained with a blocked training regime, rhyme and melody, a mode of presentation found in many children's songs (for review, see Bebout & Belke 2017).

1.3 The Present Study

The evidence reviewed in the previous section suggests that children's songs may be particularly suited as input for training the acquisition of gender subclassses in German. Like children's songs, picture books may present a means to optimize the linguistic input such that learners can pick up on relevant syntactic cues. For instance, when novel actors or entities feature in a story, they are typically referred to with an indefinite NP. Once given, they are referred to by a definite NP with the same head or a pronoun. This way, syntactic cues from multiple paradigms are presented in a grouped fashion (for text examples, see Kauffeldt et al. 2014; von Lehmden et al. 2017). Like singing songs, reading picture books can be implemented in everyday life, and it has been shown to generally support early literacy acquisition (Müller 2012).

In the present study, we investigated whether presenting a noun in (the artificial equivalents of) indefinite and definite noun phrases with inflected prenominal adjectives ([ein kleiner Mann]/[der kleine Mann])MASC.SG.NOM ('a/the little man') facilitated the acquisition of gender-like subclasses compared to artificial equivalents to phrases like ([ein kleinØ Mann]/[der kleinØ Mann]) MASC.SG.NOM. Note that in German, definite articles, when presented on their own, without reference to other syntactic cues, are ambiguous regarding the gender of the noun they accompany. For instance, *der* can be both masculine (nominative, e.g. *der Mann*, 'the man') but also, albeit more rarely, feminine (dative, *der Frau*, 'to the woman'). The same holds for indefinite articles, which are even more ambiguous. Adding an inflected prenominal adjective can help disambiguate these cues ([der klein-e Mann]: der + -e vs. [der klein-en Frau]: der + -en). In order to assess whether disambiguating the cues to a noun's gender in this way can facilitate gender acquisition, we constructed two versions of an artificial language based on the artificial language devised in Bebout & Belke (2017). The versions were largely identical except that in one version all nouns occurred with inflected prenominal adjectives, while in the other they occurred with prenominal articles only and the adjectives were integrated in the input without inflectional variation. We used a fully implicit training regime.

2. Method

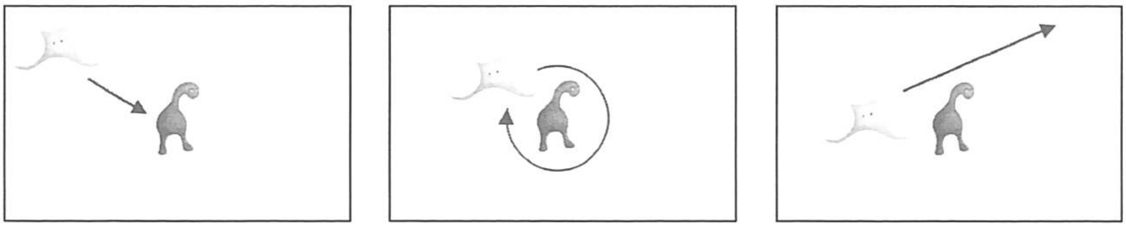
2.1 Participants

72 children from daycare centers in the Ruhr area were signed up by their parents. 24 did not complete the study for various reasons, 7 more were

excluded because they had difficulties in understanding or completing the tasks. The final sample included the data of 41 children (inflected condition: 11 girls and 10 boys aged 4;8 to 6;8 years, *M* = 5;9 years, *SD* = 6.57 months; non-inflected condition: 12 girls and 8 boys aged 4;7 to 6;8 years, *M* = 5;8 years, *SD* = 6.33 months).

2.2 Materials

The materials were adapted from Bebout & Belke (2017). At the heart of the materials and the language, there were 21 fantasy creatures, coloured shapes with little eyes whose names were pseudowords in German (see Fig. 1 for an example). 20 of these creatures interacted with another, key creature named Tika by moving towards, around or away from Tika. We introduced the study to the children by explaining to them that their task was to help Tika, a creature from planet Schwupp, learn the language of a group of creatures from another planet so as to play and dance with them. The pictures of the alien creatures had been created by Claudia Kirschke. For the purpose of the current study, some of the creatures were given a different coloured (various shades of yellow, red, and green).



to-locative by-locative from-locative

Figure 1. The three movements associated with the movement markers of the artificial language.

The lexicon of the artificial language consisted of the 21 creature names, four movement markers (ano, ino, pim, pif), and four additional markers used in the inflected training condition only (-i, -o, -am) (see Appendix A). In the following, we will refer to the markers as movement markers and adjective inflections, respectively. In addition, three colour words (gelb, grün, rot) were included in the language; to ease processing load, we used existing colour words. These colour words functioned as bearers of the adjective inflections in the inflected training group but were otherwise not relevant for training.

The names of the 20 fantasy creatures were evenly assigned to two gender-like subclassses. Per subclass, 6 creatures and their names were used in training. Of the remaining four items per subclass, two were introduced in training session 1 and revised regularly but were not involved in language training (novel-old items). The remaining ones were introduced in the very last session (novel-new items). Novel items were used in the generalization test (Test 3, see below).

	Gender-like subclass 1			Gender-like subclass 2		
Movement	Adjective	Adjective inflection	Movement marker	Adjective	Adjective inflection	Movement marker
I	gelb	-o	pim	gelb	-i	pif
II	gelb	-i	ano	gelb	-i	ino
III	gelb	-o	ino	gelb	-am	ano

Table 1. Gender-like and case-like movement markers and adjective inflections used in the artificial language.

Each subclass was associated with a distinct paradigm of grammatical markers coding whether a creature moved towards, around, or away from Tika. We will refer to these as movements I to III. Table 1 presents the movement markers and adjective inflections for movements I to III in each subclass. The first two movement markers in subclass 2 and their (non-)overlap were modelled after the German indefinite and definite articles in the nominative case for feminine nouns and the third one after the definite feminine article for the dative case (eine, die, der → pif, ano, ino). The same held for the adjective inflections, including the syncretism in the first two forms (-e, -e, -en → -i, -i, -am). Similarly, the first two markers in subclass 1 were modelled after the indefinite and definite articles in the nominative case for masculine nouns and the corresponding adjective inflections (ein + -er (pronounced [ɐ]) → pim + -o, der + -e → ano + -i). For the third movement marker and inflection, we deviated from the model (dem + -en → ino + -o) to render the present study comparable to a companion study (Belke et al., in prep.) with respect to the number of markers used and the syncretism within and across marker paradigms (see also Appendix A). Note that our goal was not to examine effects of definiteness on gender-like category acquisition, but to explore how authentic syntagmatic combinations of marker paradigms can be used as implicit cues to the gender-like subclasses of the accompanying nouns.

In total, the gender-like subclasses and movements generated six grammatical functions (two gender-like subclasses x three movements), which were filled with four movement markers (pim, ano, ino, pif) and three adjective inflections only (-o, -i, -am). This way, we introduced syncretism into the marker system, a prominent feature of nominal inflection in German. Critically, however, there was a 1:1-assignment of marker combinations to gender-movement-combinations in the inflected condition. Hence, the inflected condition featured unambiguous cues to the gender-like subclasses, but the cues were also more complex than those in the non-inflected condition, which featured fewer but ambiguous markers.

The children were trained with sentences of the form creature name – adjective (adjective inflection) – movement marker. (1) to (3) present the sentences for Lelop's movements, shown in Fig. 1, a creature of gender-like subclass 1 (G1).

(1) to:
Lelop gelb(-o) pim
Lelop.SBJ gelb(-o.G1.MI) pim.G1.MI
'Lelop moves towards Tika.'

(2) around:
Lelop gelb(-i) ano
Lelop.SBJ gelb(-i.G1.MII) ano.G1.MII
'Lelop moves around Tika.'

(3) away from:
Lelop gelb(-o) ino
Lelop.SBJ gelb(-o.G1.MIII) ino.G1.MIII
'Lelop moves away from Tika.'

2.3 Procedure

Six training and two test sessions were conducted within a span of 2.5 weeks. Sessions lasted up to 45 minutes. As detailed previously, the background story for the children taking part in training was that they were to help Tika, a creature from planet Schwupp, learn the language of creatures from another planet. Accordingly, training started with shared reading of the initial section of a picture book on Tika's adventures, and all tasks administered in training were implemented as board games involving laminated paper models of Tika and the other creatures. Presentation of sentences in training was always combined with performing the corresponding movements using the creature models and the game boards. As in Bebout & Belke (2017), the order of presentation was blocked by creature such that all three sentences pertaining to one item were presented consecutively. So, for example, all three sentences associated with the creature named "Filan" would be presented, followed by all three sentences associated with "Lelop" (see Fig. 1), and so on. Items were presented in a pseudorandomized order. Throughout all sessions, after each block of three sentences (listen-and-repeat and sentence production tasks) or after every individual sentence (sentence comprehension), the child would stick an object sticker to a game board as a reward for the alien creatures. This way, the child could see how many trials they had already completed.

During language training, the experimenter gave immediate feedback after every sentence and, if the child's answer had been incorrect, gave the correct answer. The child was then asked to repeat the correct sentence while performing the associated action. During the test sessions, the experimenter no longer gave feedback concerning whether the child's answers were correct but still offered encouragement during the tasks.

Session 1 was reserved for vocabulary training only, featuring the 12 creature names which were to appear during language training and the 4 novel-old items, the names of which were trained throughout but did not appear in any of the other training tasks. Language training proper was administered in Sessions 2 to 6. Each of these sessions started with a revision of the 16 creature names. After that, one or more training tasks were administered: a listen-and-repeat task

(Sessions 2 to 4), a comprehension task (Sessions 3 to 6) and a sentence production task (Session 3 to 6) (see Appendix B for details on the procedure).

Sessions 7 and 8 were test sessions. They both started with a revision of the vocabulary so as to establish the active vocabulary knowledge of all children. In Session 7, this was followed by a sentence production and a sentence comprehension task that both included all 12 items that had appeared during language training before (36 sentences per task). In these tasks, items were presented in a pseudo-randomized fashion such that the blocked order of presentation of movements was retained but the three movements did not apply to the same noun anymore. Instead, three different items would occur across the three movements. Children were either asked to produce the sentences to a given action or to listen to an utterance and act it out with the correct miniature figure. In Session 8, a sentence production task was used to assess the children's ability to generalize their knowledge of the artificial language to new material. It was based on the four items that had appeared during vocabulary training but not during language training (novel-old items) followed by four new items (novel-new items). Since the gender-like subclass of an item could not be inferred by its name or any external feature, the first of the three sentences (pertaining to movement I) was given as a hint by the experimenter along with the associated movement on the game board. Then the child was asked to take the novel creature, repeat the sentence while performing the movement, and then say the following two sentences (movements II and III) while also performing the associated actions.

3. Results and Discussion

Prior to analyzing the data from the tests, we assessed how well the children had mastered the vocabulary in each of the training sessions. The acquisition trajectories were very similar across conditions and subclasses (see Figure 2). Using an MCA, we reduced the vocabulary scores over sessions 2 to 6 to two dimensions, which accounted for 67.2% and 13.7% of the variance, respectively. For the analyses of the children's performance in the tests, we excluded all sentences pertaining to items whose names a child had not been able to name actively in any of the vocabulary revision phases of the training sessions (see details on exclusion criteria given with each test). We included the dimensions from the MCA in the mixed logistic regression models of the remaining data, but they were not relevant for any of these models. In the interest of brevity, we will therefore skip the analysis of the children's vocabulary knowledge and report the test results of the final mixed regression models per test.

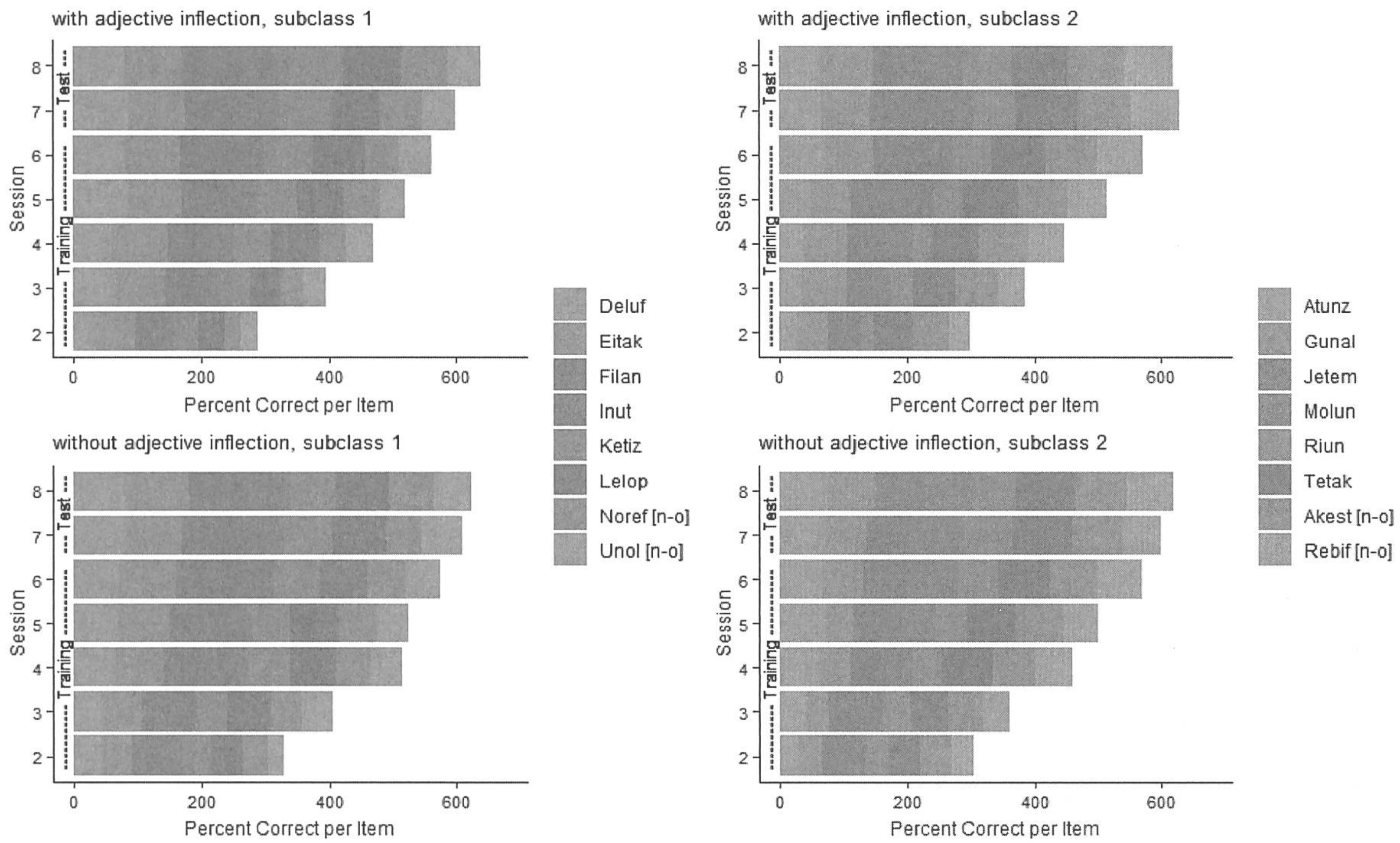


Figure 2. Average percentage of correct answers in the vocabulary revisions at the beginning of each session, broken down by gender-like subclass and training condition. Noref, Unol, Akest, and Rebif were novel-old items that were introduced in the vocabulary training in session 1 and the vocabulary revision from session 2 onwards but were withheld from all other tasks during training

3.1 Random-order Sentence Production (Old Items, Test 1)

There were two missing cases in the inflected condition due to a recording error, yielding 1474 data points in total. We excluded all sentences pertaining to items whose names the children had not been able to name actively in any of the vocabulary revision phases of the training sessions. This led to the exclusion of 139 and 147 data points in the inflected and non-inflected training conditions, respectively, corresponding to 18.4% and 20.4% of the data. The exclusion rates did not differ between conditions ($\chi^2(1) < 1$).

In 18.9% of the remaining responses, the children had asked for the creature name prior to producing their response. In more than two thirds of these cases, they had produced the correct item name during the revision of all creature names at the beginning of the session, suggesting that, in principle, they knew the name of the creature in question. This is why we decided to exclude only those trials from the analyses for which the children had not known the correct creature name during the initial vocabulary revision (inflected condition: n = 29 (4.7%), non-inflected condition: n = 31 (5.4%), $\chi^2 < 1$). One participant was excluded because they had not known any of the creature names of subclass 1 at the beginning of Session 7. Word order reversals or self-corrections were not coded as errors.

In order to compare marker production across conditions, we assessed the movement markers only that featured in the training of both groups. Correct and incorrect markers were coded as 1 and 0, respectively. On average, children gave correct responses in 35.7% and 38.9% of all cases in the inflected and non-inflected conditions, respectively, which was significantly above chance in both groups (25%, binomial test, $p < .001$). We fitted a mixed logistic regression model to the data, initially including Subclass, Condition, composite scores of the children's vocabulary knowledge, and their interactions as fixed factors and random intercepts by participants (AICc = 1485). We subsequently reduced stepwise the complexity of the model, each time choosing the model with the lowest AICc (1471) that did not differ significantly from the maximal model ($\chi^2(12) = 10.84$, $p > .5$). Table 2 presents the final model and the predicted probabilities per condition and subclass generated by the model.

General linear mixed effects model *Fixed Effects*[a]	Estimate	SE	95% CI		Wald z		Predicted probabilities
Intercept (Condition=Non-Inflected, Subclass=1)	-0.578	0.148	-0.877	-0.289	-3.91	***	36.5 (5.15)
Condition (Condition=Inflected, Subclass=1)	-0.427	0.212	-0.849	-0.010	-2.02	*	27.4 (4.73)
Subclass (Condition=Non-Inflected, Subclass=2)	0.200	0.180	-0.152	0.553	1.11		41.4 (5.44)
Subclass:Condition (Condition=Inflected, Subclass=2)	0.546	0.254	0.049	1.044	2.15	*	44.0 (5.85)

[a] Random intercept (Participants): Variance = 0.117, SE = 0.342

Table 2. General linear mixed regression of the children's responses in Test 1 (fixed and random effects) and % correct (with standard deviations) as predicted by the model.

The children in the non-inflected training group performed better than the children in the inflected training group, but this effect was restricted to subclass 1, yielding significant effects of Condition and its interaction with Subclass.

To explore these effects further, we assessed in more detail what kind of errors the children made, and which aspects of the marker paradigms were affected most. Figure 3 presents, for each training group and subclass, the percentage of cases in which a marker was used for a given movement. Target markers are shown in bold and with triangles, overgeneralizations in italics and with squares.

The markers produced most frequently by the children in the inflected training group corresponded to the correct marker paradigm in subclass 2 but not in subclass 1. In subclass 2 (top right in Figure 3), the correct markers (pif, ino, ano) were used substantially more frequently than the incorrect markers. This held in particular for movements I and III, where none of the other markers exceeded chance. This indicates that many of the children had grasped the marker paradigm of gender-like subclass 2. In subclass 1 (top left in Figure 3), the children tended to overgeneralize the markers from the paradigm for subclass 2. Note however, that they produced the correct markers almost equally often for the first and third movement, suggesting that many of the children had understood that the subclasses were associated with distinct marker paradigms. Indeed, the average frequencies with which the children in the inflected training group used each marker across subclasses differed significantly between subclasses (χ^2

(3) = 19.34, $p < .001$, cf. "Total" in Figure 3).

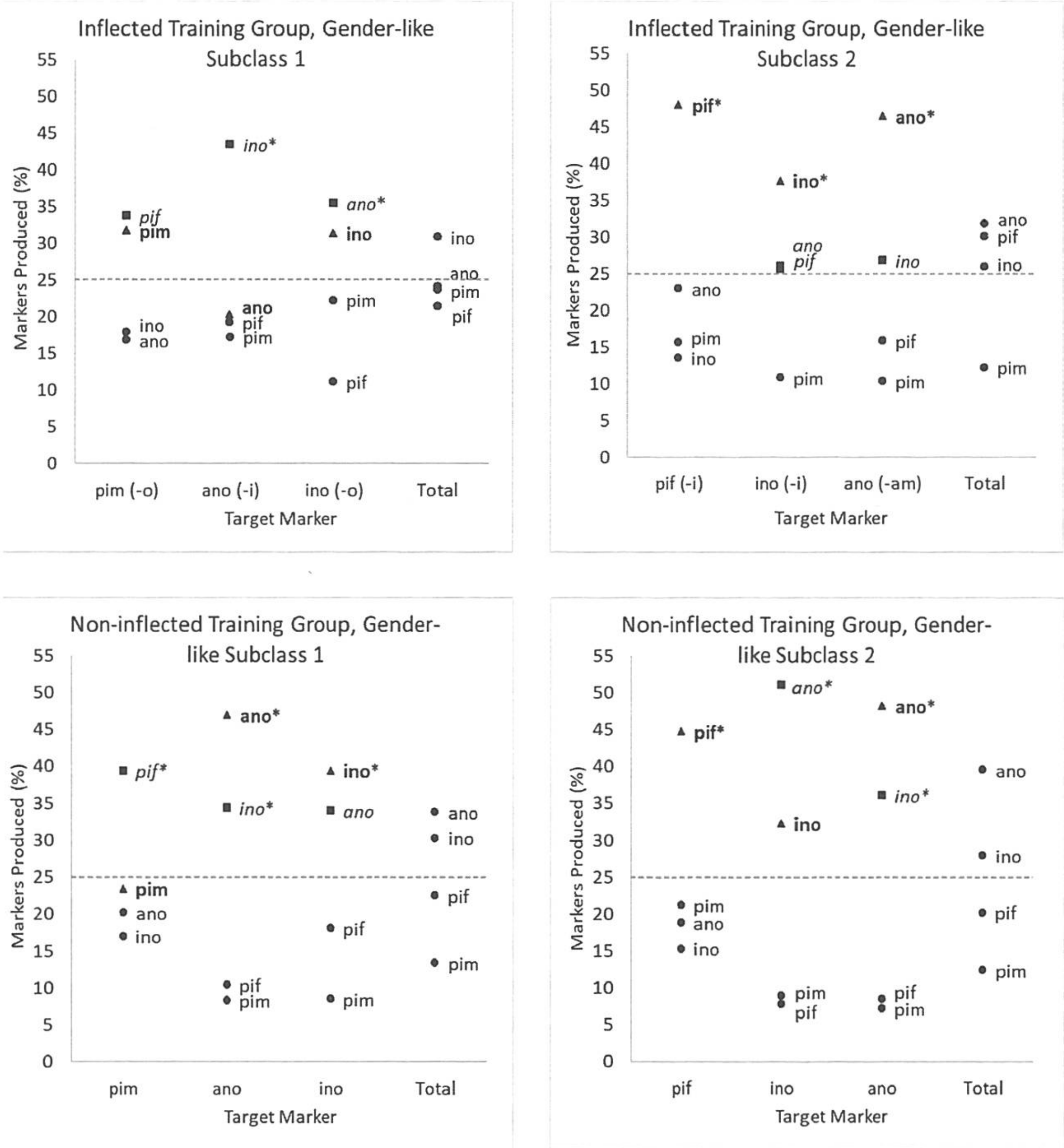


Figure 3. Percentage of cases in which a marker was produced per subclass and movement in the inflected and non-inflected training groups. Target markers are shown in bold and with triangles, overgeneralizations in italics and with squares. Dashed lines represent chance (25%); markers marked with * were produced significantly above chance.

Two aspects of the errors committed by the children in the inflected training group are particularly noteworthy: Most overgeneralization errors in subclass 1 occurred in movement II, which was associated with identical adjective inflections (-i) in both subclasses. Accordingly, most of the errors in subclass 2 pertained to this movement, too, and the movement markers the children produced instead of the target marker were those that also shared the adjective inflection -i with the target marker ino, namely ano (subclass 1) and pif (subclass 2). In contrast, rather few overgeneralization errors were committed when the target markers were pif or ano in subclass 2, arguably because they were highly distinctive (pif) or were combined with a distinctive adjective inflection (ano, with the inflection - am).

The children in the non-inflected training group showed a very different pattern of results. In subclass 1, their most frequent response pattern was pif-ano-ino, in subclass 2 pif-ano-ano, respectively (bottom row in Figure 3). Thus, in none of the subclasses the most frequently produced markers matched with all three target markers, unlike what we had seen in subclass 2 of the inflected group. In both subclasses the most frequently produced marker for movement I was pif, while other markers did not exceed chance. This indicates that the children had understood that the marker pif described movement I but the overgeneralization to subclass 1 suggests that they were unable to distinguish between the two subclasses. For movements II and III, the children produced the markers ano and ino most frequently in both subclasses, while pif and pim were rarely produced. Comparing how often the children produced each marker per subclass, we obtained no differences between gender-like subclasses ($\chi^2(3) = 1.94$, $p > .5$, see "Total" in Figure 3), suggesting that the children had not yet grasped that each subclass was associated with a different marker paradigm. The highly frequent use of the markers ano and ino in movement II and III in both subclasses and their overgeneralization are most likely due to the fact that ano and ino feature double as often than pim and pif and that they fill different functions in the marker paradigms. This led the children to keep the markers in memory but hindered them in assigning them to the markers to the correct movements in the two marker paradigms.

In sum, the results suggest that the children in the inflected training group were able to make use of those adjective inflections that were distinctive; they clearly helped them grasp that there were two different gender-like subclasses with distinctive inflectional patterns. The children in the non-inflected training group, by contrast, had not yet grasped that there were distinct subclasses associated with individual marker paradigms.

3.2 Random-order Comprehension (Test 2)

One child in the inflected training condition did not participate in Test 2 but returned for Test 3 the next day. In Test 2, this left 1440 data points from 40 children for analysis. As with Test 1, we excluded all sentences pertaining to items whose names the children had not been able to name actively in any of the vocabulary revision phases of the training sessions. This led to the exclusion of 273 and 126 data points in the inflected and non-inflected training conditions, respectively, corresponding to 17.5% and 20.4% of the data ($\chi^2(1)$ 1.81, $p > .17$).

Correct and incorrect markers were coded as 1 and 0, respectively. One-sided binomial tests showed that the children in both groups performed significantly above chance (33%, $p < .001$; inflected condition: 45.5%, non-inflected condition: 47.6%). We fitted a mixed logistic regression to the data with Condition, Subclass, composite scores of the children's vocabulary knowledge, and all interactions as fixed factors and random intercepts by participants (AICc = 1559).

General linear mixed effects model *Fixed Effects*[a]	Estimate	SE	95 % CI		Wald z		Predicted probabilities
Intercept (Condition=Non-Inflected, Subclass=1)	0.048	0.194	0.483	.0896	0.246		51.1 (12.4)
Condition (Condition=Inflected, Subclass=1)	-0.409	0.274	-0.338	0.434	-1.495		42.6 (14.3)
Subclass (Condition=Non-Inflected, Subclass=2)	-0.300	0.176	-0.958	0.134	-1.699		44.1 (12.7)
Subclass:Condition (Condition= Inflected, Subclass=2)	0.567	0.249	0.079	1.057	2.276	*	48.3 (14.3)

[a] Random intercept (Participants): Variance = 0.435, SE = 0.659

Table 3. General linear mixed regression of the children's responses in Test 2 (fixed and random effects) and predicted probabilities (with standard deviations) of the model.

Table 3 presents the final model (AICc = 1555; $\chi^2(12) = 10.84$, $p > .5$) and the percentage of correct responses as predicted by the model. There were no main effects of Condition and Subclass but a significant interaction: While the non-inflected training group performed better forsubclass 1 than for subclass 2, the reverse was true for the inflected training group. Figure 4 presents how often (in %) the children acted out which movement (m1: towards, m2: around, m3: away-from) for a given marker in each subclass. Used target movements are highlighted in green.

For movement I, the results suggest that the children of both training groups had picked up on the fact that markers pim and pif were associated with movement I (towards). For the remaining markers, the children in the inflected group performed better in subclass 2 than in subclass 1, acting out all target movements significantly above chance in subclass 2. The error pattern for subclass 1 suggests that the children in the inflected group overgeneralized the marker paradigm from subclass 2 to subclass 1. We assume that this was the case because movement III in subclass I was associated with the same movement marker (ino) as movement II in subclass 2, highlighted in blue in Figure 3, and movement II was associated with the same adjective inflection (-i) in both subclasses, highlighted in dark orange in Figure 3. Notably, the correct movements were used almost equally often as the overgeneralized ones, suggesting that there were many children who had grasped the paradigms of subclass 1 and 2.

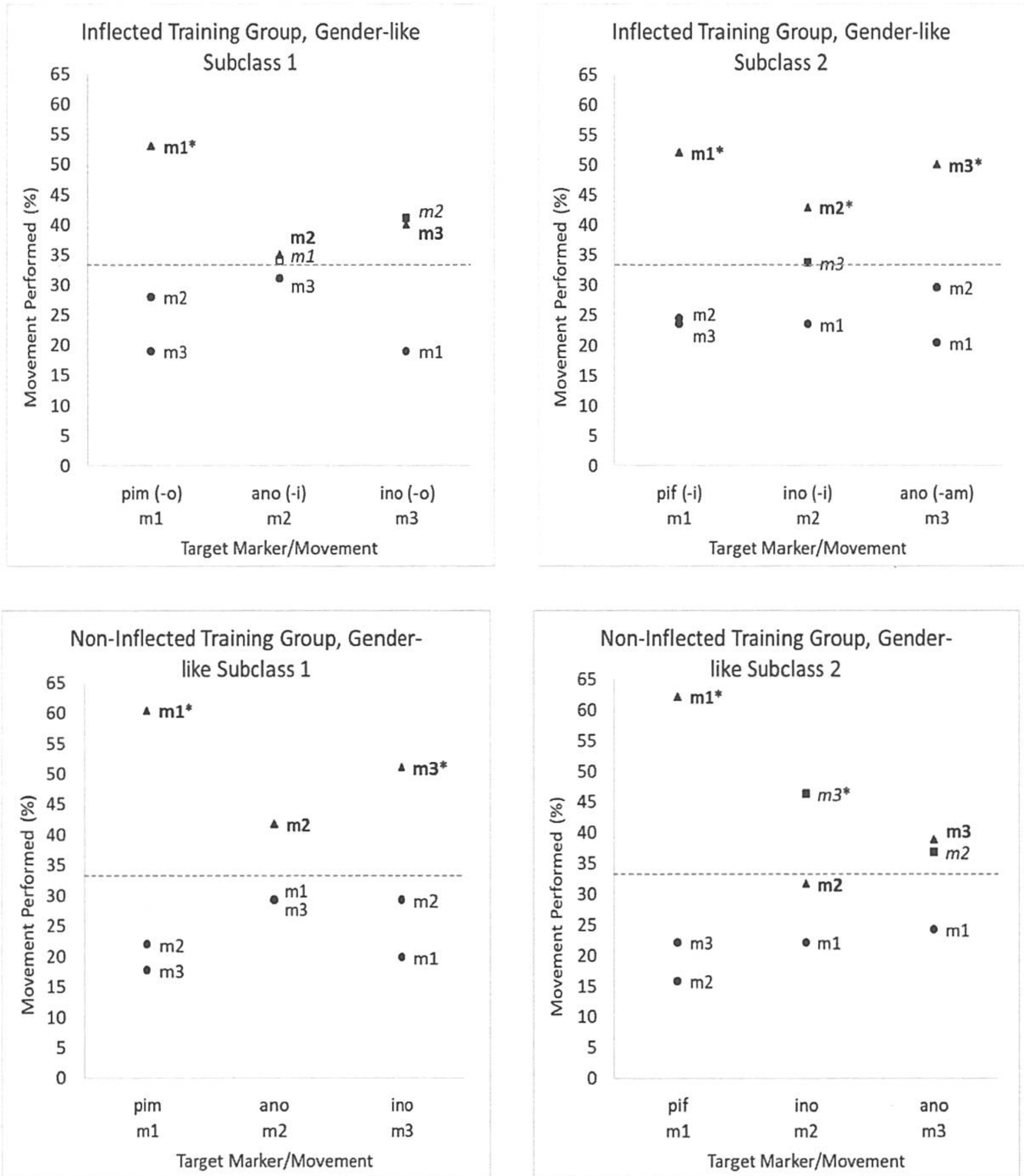


Figure 4. Frequency (in %) of the movements acted out by the children in each training group for a given marker and subclass. Target movements are shown in bold and with triangles. If produced at or above chance, overgeneralizations due to ambiguous movement markers are shown in italics and with squares with black fill colour, overgeneralizations due to ambiguous adjective inflections are shown in italics and with squares with white fill color, respectively. Dashed lines represent chance (33%); markers marked with * were produced significantly above chance.

The children in non-inflected training group showed the reverse generalization (from subclass 1 to subclass 2), accounting for why they performed more poorly in gender-like subclass 2. Their overgeneralization pattern differed from the inflected training group, however, because in movement II of subclass 2, the overgeneralization (m3) is used significantly above chance and much more frequently than the correct movement (m2). All in all, the pattern of performance seen in the comprehension task were parallel to those seen in the production task (Test 1).

3.3 Random-order Sentence Production (Novel Items, Test 3)

In Test 3, all children were given the sentence for the first action of a novel creature and were asked to produce the sentences to the remaining two actions. If the children had acquired the paradigm, they would recognize the marker in the first sentence as the first of a triple of markers and produce the two other sentences with the correct markers accordingly. We tested two novel-old and two novel-new items per gender-like subclass and child, yielding 328 pairs of sentence responses in total. As the name of the creature was provided in the initial sentence cue, we did not exclude any data points due to lack of vocabulary knowledge. Novel-old item names the children were not able to produce during the vocabulary revision at the beginning of the test session were coded as novel-new (subclass 1/2: 38.1%/16.7% and 35.0%/22.5% of the novel-old data points in the inflected and non-inflected training conditions, respectively).

There was no difference between training groups in their accuracy with which they used the markers ($\chi^2(1) < 1$, mean accuracy: 41%); this was confirmed in an independent linear mixed effects regression on the data. There were also no differences in the children's overall frequency of using each marker, irrespective of its correctness ($\chi^2(3) = 4.7$, $p = .19$). As one might expect, most erroneous responses involved combinations of ano and ino, which were the two target forms in Test 3.

In qualitative analyses, we assessed how often the children produced individual combinations of markers (Table 4), collapsed over novel-old and novel-new items, which yielded parallel patterns of results. The children in the inflected training group overgeneralized the marker combination ino-ano from subclass 2 to subclass 1 in about 31% of the cases, but the reverse did not occur. In fact, the marker combination ano-ino was produced in only 7.1% of all cases in both subclasses. Arguably, the adjective markers in this subclass were more salient, especially -am in movement III, allowing the children to extract them more effectively. By contrast, the children in the non-inflected training group overused both ino-ano and ano-ino (cf. lines 1 and 2 for the non-inflected training group). As it is not possible to infer from Table 4 which child contributed to which marker pattern in a given subclass, we assessed how often the children of each group used a marker pair in each subclass. We obtained parallel results across groups ($\chi^2(2) = 1.92$, $p > .3$): Most children used a marker pair slightly more often in one subclass than the other or equally often in both subclasses (57% and 19%

of all cases, respectively). Patterns that were clearly used more often in one subclass only occurred substantially less often (24%). This suggests that none of the groups were able to generalize the patterns acquired in training to novel items. Overall, the children's patterns of performance in the generalization test mirrors the patterns observed in Tests 1 and 2 for trained items: While the children in the inflected training group overgeneralized in one direction only, the children in the non-inflected training group overgeneralized in both directions.

Inflected Training Group					
Gender-like Subclass 1			Gender-like Subclass 2		
Responses	Freq in %	*p*	Responses	Freq in %	*p*
X ino – X ano	*31.0*		**X ino – X ano**	**28.6**	.001
X ano – X ano	14.3		X ano – X ano	15.5	
X ino – X ino	11.9		X ino – X pif	14.3	
X ano – X ino	**7.1**	n.s.	X ano – X pif	10.7	
X ino – X pif	7.1		*X ano – X ino*	*7.1*	
X pif – X ano	7.1		X ino – X ino	6.0	
Non-inflected Training Group					
Gender-like Subclass 1			Gender-like Subclass 2		
Responses	Freq in %	*p*	Responses	Freq in %	*p*
X ano – X ino	**30.0**	.001	*X ano – X ino*	*35.0*	
X ino – X ano	*21.2*		**X ino – X ano**	**22.5**	.001
X ano – X ano	15.0		X ano – X ano	10.0	
X ano – X pif	15.0		X ano – X pif	8.8	
			X ino – X ino	5.5	

Table 4. Pairs of sentence responses produced most frequently by the children of the non-inflected and the inflected training group. For each group and subclass, the marker pairs listed make up about 80% of the children's responses. The target pair is printed in bold, pairs constituting overgeneralization errors in italics. *p*-values were established using binomial tests against chance (6.25%).

4. General Discussion

The aim of the present study was to investigate whether implicit learning of gender-like subclasses in preschool children can be advanced by means of implicit teaching via input optimization. To this end, we assessed how well preschool children aged 4;6 to 6;6 acquired gender-like subclasses in an artificial language. The language and training provided syntactic cues to the grammatical subclasses of the nouns only. We compared the performance of a group of children trained with one set of markers (non-inflected condition) to that of a group trained with the same set of markers plus additional adjective inflections (inflected condition). In the inflected condition, the syntactic cues were unambiguous cues to the gender-like subclasses. However, the cues were also phonologically more complex than those in the non-inflected condition, which featured fewer but ambiguous markers.

We found that none of the two groups were able to generalize the patterns acquired during training to novel items, suggesting that the training would have had to be implemented for longer to obtain generalization effects. In tests of sentence production with trained items, the children in the inflected training group performed more poorly than the children in the non-inflected training group, suggesting that the additional hints were a hindrance to the acquisition of the gender-like subclasses. However, the effect of training condition was restricted to gender-like subclass 1 only. In both, sentence production and sentence comprehension (Test 1 and 2) the analyses of the marker errors the children committed suggest that many children in the inflected training group had picked up on the fact that there were two different gender-like subclasses and that the marker paradigms differed between subclasses. While they had acquired the paradigm for subclass 2, they had not yet acquired the correct paradigm for subclass 1. However, the fact that they used the correct markers equally often as the overgeneralized ones indicates that many of the children had already grasped the marker paradigm of subclass 1 as well. We hypothesize that the children in the inflected training group were able to make use of distinctive adjective inflections, chiefly -am in subclass 2, which helped them pick up on the relevance of the marker combinations.

The children in the non-inflected training group used the markers equally often in both subclasses, except for marker ano in movement III, which they used more often in subclass 2. In none of the subclasses the most frequently produced markers matched with all three target markers, unlike what we had seen in subclass 2 of the inflected group. We take this to suggest that they had not yet grasped that there were distinct gender-like subclasses associated with individual marker paradigms.

We conclude that combinations of marker paradigms can facilitate gender-like subclass acquisition but that learners must be given sufficient time to acquire the relevant marker combinations. Although the input of the inflected group imposed more load on phonological STM and phonological working memory the error quality of used grammatical markers indicate that in contrast to the non-inflected group many of the children had understood the existence of two subclasses with different marker paradigms and not only overgeneralized one preferred marker paradigm. Thus, even if both groups of condition would benefit from more time to acquire the paradigms, the errors committed suggest that the inflected group almost reached solid acquisition of both marker paradigms.

Our findings are relevant beyond oral language acquisition in that grammatical gender is key to anaphoric relations in written, literate texts and hence plays a central role in literacy acquisition.

REFERENCES

Bebout, J. & Belke, E. (2017). Language play facilitates language learning: Optimizing the input for gender-like category induction. *Cognitive Research: Principles and Implications, 2(1),* 11.

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 (pp. 174-200). Baltmannsweiler: Schneider Hohengehren.

Belke, E., Kuba, S., & Braun, P. M. (2022, September 7-9). Optimizing the presentation of the input facilitates the acquisition of gender-like subclasses in preschool children. Poster presentation, AMLaP, York, UK.

Brooks, P. J., Braine, M. D. S., Catalano, L., & Brody, R. E. (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.

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

Culbertson, J. & Schuler, K. (2019). Artificial language learning in children. *Annual Review of Linguistics, 5,* 353-373.

Gathercole, S. E. (2007). Workin memory and language. In M. G. Gaskell (ed.), *The Oxford handbook of psycholinguistics* (pp. 757-769). Oxford: OUP.

Jeuk, S. (2008). "Der Katze jagt den Vogel". Aspekte des Genuserwerbs im Grundschulalter. In B. Ahrenholz (ed.), *Zweitspracherwerb: Diagnosen, Verläufe, Voraussetzungen* (pp. 135-150). Freiburg: Fillibach.

Kauffeldt, L., Kirschke, C., Bebout, J., Frieg, H., Belke, E., Hoffmann, R., & Belke, G. (2014). *Dschungeltanz und Monsterboogie. Singen und Spielen mit Sprache* [Jungle dance and monster boogie. Singing and playing with language]. Baltmannsweiler: Schneider Hohengehren.

Köpcke, K. M. & Zubin, D. A. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte, 93,* 26-50.

Müller, C. (2012). *Kindliche Erzählfähigkeiten und (schrift-)sprachsozialisatorische Einflüsse in der Familie. Eine longitudinale Einzelfallstudie mit ein- und mehrsprachigen (Vor-)Schulkindern.* Baltmannsweiler: Schneider Hohengehren.

Ruberg, T. (2015). Diagnostische Aspekte des Genuserwerbs ein- und mehrsprachiger Kinder. *Forschung Sprache, 2,* 22-40.

Savage, C., Lieven, E., Theakston, A. & Tomasello, M. (2006). Structural priming in language acquisition: The persistence of lexical and structural priming in 4-year-olds. *Language Learning and Development, 2,* 27-49.

Schröder, A., Kauschke, C. & De Bleser, R. (2003). Messungen des Erwerbsalters für konkrete Nomina. *Neurolinguistik, 17,* 83-114.

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,* 445-471.

Taraban, R. (2004). Drawing learners' attention to syntactic context aids gender-like category induction. *Journal of Memory and Language, 51,* 202-216.

Ullman, M. T., & Pierpont, E. I. (2005). Specific Language Impairment is not specific to language: The Procedural Deficit Hypothesis. *Cortex, 41,* 399-433.

Ulrich, T. (2017). *Grammatikerwerb und grammatische Störungen im Kindesalter. Ergebnisse des Forschungsprojekts GED 4-9 und ihre Implikationen für sprachdiagnostische und -therapeutische Methoden* Unpublished habilitation thesis. Kölner UniversitätsPublikationsServer. https://kups.ub.uni-koeln.de/9011/

Verhagen, J. & Lesemanm P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141,* 65-82.

von Lehmden, F., Müller-Brauers, C., Belke, E., Belke, G., & Lehmann, B. (2017). *Immer anders. Bilderbuch zum impliziten Grammatiklernen.* [Always different. A picture book for learning grammar implicitly]. Baltmannsweiler: Schneider Verlag Hohengehren.

Wegener, H. (1995). *Die Nominalflexion des Deutschen – verstanden als Lerngegenstand.* Tübingen: Niemeyer.

Appendix A: The Artificial Language

Table A1 provides a list of all creature nouns, except Tika, i. e. the names of the creatures interacting with Tika. All creatures were introduced to the children as inhabitants of a far-away planet, i. e. as living beings. We used a subset of the pseudonouns and images used by Belke & Bebout (2017) for our study, including only two instead of three subclasses. The nouns from Belke & Bebout (2017) were pseudowords of German and did not include phonological cues to one gender subcategory. Their written forms consisted of either four or five letters, and they were pronounced as disyllabic words consisting of four to six phonemes.

Gender-like subclass 1		Gender-like subclass 2	
Deluf	Noref (novel-old)	Atunz	Akest (novel-old)
Eitak	Unol (novel-old)	Gunal	Rebif (novel-old)
Filan	Nelun (novel-new)	Jetem	Abun (novel-new)
Inut	Elwan (novel-new)	Molun	Runot (novel-new)
Ketiz		Riun	
Lelop		Tetak	

Table A1. Nouns of the artificial language for both gender-like subclasses.

To ensure that neither the nouns nor the creatures' appearances included any cues to subcategory, Belke & Bebout (2017) had conducted a norming study. Adult participants were presented images depicting the creatures and were asked to indicate whether they considered each creature male, female, or neuter, by choosing the definite article they associated with the creature. The same participants listened to audio recordings of the nouns and were asked to write down each pseudoword and to select the definite article they considered most fitting. Creatures and nouns which were associated with one gender category in more than 50% of cases were assigned evenly to the three subclasses of the artificial language, while all other nouns were randomly assigned to one of the subclasses.

Consequently, it was impossible for our participants to infer the nouns' subclass using characteristics like the number of syllables, phonological cues, or the creatures' appearance. Granted, this did not exclude the possibility of participants searching for cues. Nonetheless, this attempt would not have led to correct classifications, as there were no valid cues to the nouns' subcategories.

Table A2 presents the marker paradigms used in the present study. In both subclasses, the movement markers and the adjective inflections were modelled after the indefinite and definite articles in the nominative case and the definite articles in the dative case for masculine and feminine nouns, respectively. For the third movement marker and inflection in subcategory 1, we deviated from the model to render the present study comparable to a companion study (Kuba et al. in prep.) with respect to the number of markers used and the syncretism within and across marker paradigms.

German masculine gender		Artificial language subclass 1		German feminine Gender		Artificial language subclass 2	
Article	Adjective inflection	Movement marker	Adjective inflection	Article	Adjective inflection	Movement marker	Adjective inflection
Ein	-er	pim	-o	eine	-e	pif	-i
Der	-e	ano	-i	die	-e	ano	-i
Dem	-en	ino	-o	der	-en	ino	-am

Table A2. German marker paradigms and the paradigms used in the artificial language.

Appendix B: Procedural Details

B.1 Ethics and Consent

Ethical approval was obtained from the ethical board of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS, application no. 16-04). Training and test were conducted at the children's daycare centers. Their parents or caregivers were informed about the general purpose of the study beforehand and consented to their children's participation. After the end of their children's participation, they were debriefed about the details of the experiment's goals. In case a child did not want to participate, struggled during language training, or appeared to feel uneasy during the sessions, training was not continued. Each child received a colouring book as a thank-you gift.

B.2 Detailed Description of the Tasks Used in the Training Sessions

B.2.1 Session 1 (Vocabulary Training)

Session 1 was conducted as a group activity for up to four children at a time. The children first met the experimenter and after reading the introductory chapter of the background story on Tika, a board game was played. It was designed to present the 12 creature names which were to appear during language training and the 4 novel-old items, the names of which were trained throughout but that did not feature in training. As part of the game, the experimenter gave each creature's name, which the children were asked to repeat. Following the game, Tika's wish to learn the alien language was introduced and each child received a memory game to take home. The parents were instructed to help their children learn the creature names while playing different games using the memory game cards.

B.2.2 Sessions 2 to 6 (Language Training)

Between session 1 and session 2, there was a break of about seven days to give parents and children sufficient time to practice the creature names. From session 2 onwards, all sessions were conducted with one child at a time. At the beginning of each session, the child was asked to name each of the 16 items which had been introduced during vocabulary training. For creatures the child could not name, the experimenter would give the name and ask the child to find the respective creature (e.g. the experimenter would ask, "Who is Lelop?"). The creatures the child could not name actively or not at all were revised using a game during which the child had to indicate which creature was which multiple times.

Language training started with a listen-and-repeat-task in which all 36 sentences were introduced (each of the 12 training items was associated with three sentences). The experimenter took the first creature and performed the three movements with the creature models on the game board while saying the three associated sentences. Subsequently, the child was asked to take the creature models and, after the experimenter had said the first sentence again, repeat the sentence while performing the movements themselves. After feedback was given, the experimenter said the next sentence, the child repeated the sentence and performed the movement, and feedback was given.

The tasks in sessions 3 and 4 were identical except that two different sets of items were used for the listen-and-repeat task and the sentence production task: Only one half of the items (three items each from gender-like subclass I and II) were used in session 3, the other half was used in session 4. As in session 2, the sessions started with a review of the creature names the child already knew and a revision of the items the child was not able to name. The procedure for the following listen-and-repeat-task was the same as in session 2 but included only 18 sentences this time (6 items with 3 sentences each). After that, a sentence comprehension task was administered, featuring all 12 items. For each item, the experimenter said a sentence and asked the child to act out the movement with the creature model on the game board while repeating the sentence. The last task of the session was a sentence production task, in which the same 6 items as in the listen-and-repeat task were used. The child was asked to perform the three movements for each item and produce the associated sentences to the best of their ability. If the child did not remember the creature's name correctly, the experimenter provided the name.

Like Sessions 3 and 4, Sessions 5 and 6 included the same types of tasks using different sets of items. Again, the sessions started with a review of the items the child could name followed by a revision task with the items the child was not able to name actively. Subsequently, a comprehension task using all 12 items (one sentence each) and a production task using 6 items with all three sentences were conducted. The procedures for both tasks were the same as in sessions 3 and 4 and one half of the items was used during the production task in session 5, the other half was used in the same task in session 6.