Asymmetries in the Acquisition of Subject-Verb Agreement in Dutch? Evidence from Comprehension and Production

Abstract:
Across languages, children do not comprehend 3SG/3PL subject-verb agreement before age five, despite early mastery in spontaneous speech. This study investigates subject-verb agreement in a language hitherto not studied in this respect, namely Dutch. We examine if (1) Dutch two- and three-year-olds comprehend subject-verb agreement and (2) a comprehension-production asymmetry still exists if task materials are kept constant across domains. Dutch-speaking two- and three-year-olds completed a comprehension (picture selection) and production (sentence completion) task testing 3SG and 3PL. In comprehension, both groups performed above chance on 3PL but not on 3SG. In production, accuracy on 3PL and 3SG was slightly higher than in comprehension. Comprehension and production were moderately correlated. These results show that comprehension is earlier in Dutch than in previously investigated languages, but only for 3PL, suggesting that phonological salience and cue reliability are important. They also show an asymmetry between comprehension and production, albeit much smaller than assumed in previous studies.
Asymmetries in the Acquisition of Subject-Verb Agreement in Dutch?
Evidence from Comprehension and Production

Across languages, it takes until the age of five before children use subject-verb agreement as a cue to resolve singular-plural contrasts in comprehension. Specifically, studies on English-, Spanish-, Xhosa- and German-speaking children show that children younger than five do not perform above chance when matching third person singular (3SG) and third person plural (3PL) sentences to pictures of one versus multiple referents (Johnson, de Villiers, & Seymour, 2005 for English; Pérez-Leroux, 2005 for Spanish; Gxilishe, Smouse, Xhalisa, & de Villiers, 2009 for Xhosa; Brandt-Kobele & Höhle, 2010 for German). These findings stand in sharp contrast with observations from spontaneous speech that three-year-old children reach high accuracy on subject-verb agreement marking in production (Brown, 1973 for English; Montrul, 2004 for Spanish; de Villiers & Gxilishe, 2008 for Xhosa; Clahsen, 1986 for German; Leonard et al., 1992 for Italian). In this study, we investigate the comprehension of subject-verb agreement in a language hitherto not studied in this respect, namely Dutch. We target two questions. First, we ask if Dutch-speaking two- and three-year old children are able to comprehend subject-verb agreement. Second, we ask if there is evidence for an asymmetry between comprehension and production in the acquisition of subject-verb agreement in Dutch.

Cross-Linguistic Differences in the Comprehension of Subject-Verb Agreement

Previous studies have investigated when children grasp the distinction between singular and plural agreement in various languages. For English, Johnson et al. (2005) conducted a picture-selection experiment, presenting children with sentences in which the only cue for either a singular or plural meaning was the agreement suffix on the verb (e.g., *The duck swims on the pond* vs. *The ducks swim on the pond*). In these sentences, the plural
morpheme on the noun is masked due to co-articulation effects between the noun and the verb, so the only disambiguation cue is the presence or absence of ‘-s’ on the verb indicating either a 3SG or a 3PL reading. Johnson et al. found that five- and six-year-old children performed above chance on picture-selection, but three- and four-year-old children did not. Accuracy was higher on 3SG than on 3PL in all age groups. To explain children’s poor overall performance, the authors propose that the English inflectional system is too weak to provide a cue.

Pérez-Leroux (2005) replicated Johnson et al.’s experiment in Spanish. Spanish has a richer inflectional paradigm compared to English and has pro-drop. Despite these differences between Spanish and English, Pérez-Leroux found very similar results: Performance was above chance in children aged between 4;8 and 6;6, but not in children between three- and 4;5-years. This is further evidence that the comprehension of number agreement is late, even in a language with a rich inflectional paradigm like Spanish in which subjects are often covert. A difference between English- and Spanish-speaking children was also found: Whereas English children performed better on 3SG than on 3PL, Spanish children performed better on 3PL than on 3SG. According to Pérez-Leroux, this indicates that children acquire overt before null morphemes.

Late comprehension was also found in German, a language that expresses both 3SG and 3PL with overt morphemes. Brandt-Kobele and Höhle (2010) tested three- to four-year-old German-speaking children on a picture selection task similar to the task used in Johnson et al. and Pérez-Leroux. Test sentences contained the personal pronoun sie that is homophonous between a 3SG (feminine) and 3PL meaning, thereby leaving verbal number agreement as the only disambiguation cue (e.g., *Sie fütter-t einen Hund* ‘She feeds a dog’ vs. *Sie fütter-n einen Hund* ‘They feed a dog’). Brandt-Kobele and Höhle found that German
three- to four-year olds did not perform above chance level in a picture selection task, on neither of the two morphemes.

This review of studies shows that, cross-linguistically, children younger than five are unable to use number agreement as a cue in comprehension. This holds for English and Spanish in which either 3SG or 3PL is expressed with a null morpheme as well as German in which both morphemes are overt. It was further replicated for Xhosa, a language in which both morphemes are overt and expressed through prefixation (Gxilishe et al., 2009).

An exception to these studies showing late comprehension of number agreement is a study on French by Legendre, Barrière, Goyet, and Nazzi (2010). These authors tested French children on their comprehension of the singular-plural contrast in a preferential looking and a (manual) selection experiment. In the selection experiment, 24- and 30-month-old French children were presented with sentences containing the phrases *Il embrasse* ‘He kisses’ and *Ils embrassent* ‘They kiss’. The crucial contrast between these phrases involves the pronunciation of the subject pronouns. In French, the 3SG and 3PL subject pronouns are normally pronounced in the same way (i.e., /il/), but when the following verb is vowel-initial, the final consonant of the 3PL pronoun (/z/) is pronounced (a phenomenon called ‘liaison’). Following others, Legendre et al. assume that /z/ is situated in the verb’s onset and thus should be analyzed as an agreement marker. Their aim was to investigate if young French children are sensitive to this singular-plural contrast in comprehension. They found that 30-month-olds – but not 24-month-olds – performed above chance on both 3SG and 3PL.

To explain French children’s precocious comprehension, Legendre, Culbertson, Zaroukian, Hsin, Barrière and Nazzi (2013) propose an account in terms of morphophonological differences across languages. Specifically, they argue that two factors play a role: perceptual salience and cue reliability of agreement morphemes. As for perceptual salience, Legendre et al. (2013) cite work on speech perception suggesting that /z/ in French
liaison is more salient to detect than Spanish /n/, that is in turn more salient than English /s/.

As for cue reliability, or the extent to which there is a transparent, one-to-one relationship between a morpheme and its function, several researchers have proposed this as a factor explaining both within and across language differences in the acquisition of morphemes (cf. Brown, 1973; MacWhinney et al., 1984). Cue reliability can explain why number agreement is later in English and Spanish than in French: English ‘-s’ is opaque, because the same morpheme marks plural on nouns, and is also used for the possessive. Spanish ‘-n’ does not provide a very reliable cue either, as it marks 2PL but not 1PL and often occurs in nouns and verbs. French /z/, in contrast, is very reliable, because it is extremely rare as an initial consonant in French words, and marks plural on both nouns and verbs (Legendre et al., 2013).

**Comprehension-Production Asymmetry**

Production studies show that, across languages, children acquire subject-verb agreement at a relatively young age in spontaneous speech. For English, Brown (1973) found 90% correct use of 3SG ‘-s’ in obligatory contexts between 26 and 46 months of age (Brown, 1973). In Spanish, number morphemes emerge very early in children’s spontaneous speech, around 1;6-1;7 (Montrul, 2004) and may be productive in children as young as 1;10-2;2 (Gathercole, Sebastián, & Soto, 2002). German-speaking children reach 90% accuracy on 3SG and 3PL in spontaneous speech at three years (Clahsen, 1986), or even earlier, around age two (Poeppel & Wexler, 1993).

This high production accuracy at early stages of acquisition is surprising in light of children’s late comprehension of subject-verb agreement. Yet, it may be premature to conclude that there is an asymmetry between comprehension and production in number agreement acquisition: None of the previous studies have directly compared comprehension and production in the same children. Rather, previous claims are based on comparisons between comprehension data, collected through picture selection, and production data from
spontaneous speech. Spontaneous speech is likely to overestimate children’s agreement marking abilities, however, as it typically contains familiar and high-frequent verbs on which subject-verb agreement emerges first (Rubino & Pine, 1998; Wilson, 2003). Evidence that spontaneous speech data may present an overestimate comes from Brandt-Kobele and Höhle’s (2010) study on German children. These authors collected parental report data on children’s production of 3SG and 3PL morphemes and found that in their sample of three- to four-year-olds, 72% of the children showed productive use of 3SG and only 48% showed productive use of 3PL, questioning the idea of an asymmetrical development between comprehension and production. In a preferential looking study using the same materials as the picture selection experiment, these authors also found that German three- and four-year-olds looked longer to the matching pictures in both a 3SG and 3PL condition. This suggests that they had some knowledge of subject-verb agreement that, for some reason, did not show up in their selection of the pictures through pointing. The authors argue that comprehension-production asymmetries may result from diverging task demands rather than from genuine differences in children’s comprehension or production abilities. Specifically, spontaneous speech may overestimate children’s production abilities as compared to production tasks or parental report, while comprehension experiments using picture selection may underestimate children’s comprehension abilities as compared to looking experiments in which no manual (pointing) responses are required.

This Study: Subject-Verb Agreement in Dutch

In this study, we examine the comprehension and production of subject-verb agreement in Dutch. Dutch is different from English, Spanish and French in that both 3SG and 3PL are marked with an overt agreement suffix. It is similar to German, except that the 3PL morpheme is more salient in Dutch, because the Dutch plural morpheme adds an additional syllable to the verb stem whereas the German plural morpheme does not (e.g.,
Dutch voer-en vs. German fütter-n, ‘feed’). The sentences in (1) and (2) illustrate Dutch 3SG and 3PL agreement. An overview of the whole paradigm is given in Table 1.

(1) Het meisje drink-t ‘The girl drink-3SG’

(2) De meisjes drink-en ‘The girls drink-3PL’

[Insert Table 1 about here]

In terms of phonological salience, the 3PL form is more salient, due to its extra-syllabic character, than 3SG ‘-t’. Cue reliability is also stronger for 3PL than for 3SG. First, 3PL is used to mark all plural forms in the paradigm, whereas 3SG is also used for 2SG but not for 1SG. Second, the plural suffix ‘-en’ also marks plurality on nouns (e.g., fiets-en ‘bicycles’), but there is no such additional ‘singularity function’ for 3SG. For both morphemes, additional functions are found in Dutch. The ‘-en’ suffix occurs on infinitives (e.g., surf-en ‘surf’) as well as strong participles (e.g., gelop-en ‘walked’), and the ‘-t’ suffix also occurs on regular past participle forms such as (e.g., gesurf-t ‘surf ed’). In this respect, Dutch differs from French in which /z/ is only used to express a plural meaning, and does not have other functions.

Dutch-speaking children start to use subject-verb agreement more productively around age 2;6 in spontaneous production; that is, from this age they contrast different agreement markings and use agreement marking with a wider variety of verb types (Blom & Wijnen, 2013). Evidence from elicited production shows highly accurate and productive use of 3SG and 3PL at age three (Polišenská, 2010). Specifically, Polišenská (2010) tested twelve Dutch three-year-olds in an elicitation task in which children were asked to finish prompt sentences.
by the experimenter. She found that accuracy scores were 93% and 100% on 3SG and 3PL marking of existing verbs and 100% and 93% on 3SG and 3PL marking of nonsense verbs. This high performance on both existing and nonsense verbs indicates that children’s production of 3SG and 3PL agreement suffixes is productive at three years. However, in this study, only two existing verbs (tekenen ‘draw’ and drinken ‘drink’) and two nonsense verbs were included, and a relatively small sample of three-year-olds participated ($N = 12$). Previous studies have shown that children’s first productions of subject-verb agreement are often lexically specific and subject to individual differences (Gathercole et al., 2002; Wilson, 2003). So, a larger sample of Dutch children and a larger set of different verbs are needed to investigate Dutch children’s production of subject-verb agreement. In addition, in order to investigate which morpheme is acquired first (3SG or 3PL), younger children need to be tested than in Polišenská (2010), to avoid ceiling effects.

To date, no studies have investigated when Dutch children develop an understanding of verbal number agreement, and none of the previous cross-linguistic studies on the comprehension of number agreement have directly compared children’s comprehension and production of verbal number agreement using the same task materials. The aim of this study is twofold. First, we investigate Dutch children’s comprehension of 3SG and 3PL morphemes, using a similar picture selection task as in the above-described studies on English, Spanish, German, and French. Participants in our study were two- and three-year-old Dutch-speaking children. We predict that, due to the high phonological salience and strong cue reliability of the Dutch 3PL morpheme, Dutch children will show earlier comprehension of this morpheme than of 3SG. Moreover, given that 3PL has a higher salience and cue reliability as compared to agreement morphemes in English, Spanish, and German, but not

---

1 In Polišenská (2010), a head-turn preference was also conducted among Dutch 18- and 19-month-olds, showing that Dutch infants were sensitive to 3SG agreement errors but not 3PL errors, as suggested by increased listening times. Different task materials were used than in the production task, however. Also, infants’ looking behavior is indicative of early sensitivity to (un)grammaticality, and not necessarily comprehension.
French, we predict that comprehension in Dutch is earlier than in English, Spanish and German, but not necessarily earlier than in French. The second aim of our study is to compare comprehension and production in the same children, using the same stimuli. In so doing, we try to minimize the possibility of finding comprehension-production asymmetries that are due to diverging methodologies (cf. Brandt-Kobele & Höhle, 2010). We consider three possible outcomes: that comprehension follows production, that production follows comprehension, or that the two appear to be roughly contemporaneous.

Method

Participants

Participants were 29 two-year-old Dutch monolingual children with a mean age of 31 months (SD = 3, range = 24 – 35) and 41 three-year-old Dutch monolingual children with a mean age of 42 months (SD = 3, range = 36 – 47). These children were selected from a larger sample of 103 children. Children were excluded from this sample if they suffered from hearing impairment (1), had an elevated familial risk at language impairment or dyslexia (7), were not monolingual Dutch (2), or if they had not completed the whole comprehension task or at least half of the production task (20). The group of two-year-olds contained 16 boys and 13 girls. The group of three-year-olds contained 19 boys and 22 girls. In both age groups, the majority of children came from high SES families, defined as families in which one or both parents had completed higher education (60% of two-year-olds; 64% of three-year-olds). Children were recruited from four different geographic areas in the Netherlands. There is much variation in verbal inflection in Dutch dialects (Bennis & MacLean, 2006), but this was not an important factor in the current study, since all children were learning standard Dutch.

Materials

2 The inclusion criterion in the comprehension task was set at 100%, as there was a fairly high chance level in this task. In the production task, there was no such high chance level, so 50% completion was used as the criterion in this task to retain sufficient numbers for the analyses.
Comprehension task. Comprehension was tested through a picture selection task in which children selected one out of two pictures after listening to a sentence. The experiment was modeled after the experiment reported in Brandt-Kobele and Höhle (2010) in that all experimental items contained the Dutch pronoun *ze* ‘she/they’ that is ambiguous between a (feminine) 3SG and 3PL meaning. As such, the agreement morpheme on the verb provided the only cue for disambiguating between a singular and plural meaning of the sentences (see Figure 1 for an example item). The task was presented on a laptop and the presentation of the stimuli was controlled through the experimental software Eprime 2.0 (Psychology Software Tools, Pittsburgh, PA). Sentences had been prerecorded by a female speech therapist in a child-friendly voice. The pictures were all high quality photographs of real situations, acted out with playmobil figures. Recall that the pronoun ‘ze’ is ambiguous between third person singular female and third person plural and thus, to ensure ambiguity of the subject pronoun, all characters had female features.

The task contained twelve experimental items, divided over a plural and a singular condition, and six filler items. The verbs used in the experimental items were the following: *aaien* ‘stroke’, *bouwen* ‘build’, *dragen* ‘carry’, *duwen* ‘push’, *gooien* ‘throw’, *kammen* ‘comb’, *kijken* ‘watch’, *klimmen* ‘climb’, *lezen* ‘read’, *lopen* ‘walk’, *zoeken* ‘look for’, and *slapen* ‘sleep’. Previous research on Dutch-speaking children has shown effects of coda consonants on 3SG production such that ‘-t’ is produced more accurately on verbs of which the stem ends in a sonorant rather than non-sonorant (Blom, Vasić, & De Jong, 2014).

3 Unlike in (colloquial) English, where the use of ‘they’ as a marker of 3SG (next to 3PL) is a recent development, Dutch ‘ze’ (or ‘zij’) has been used as a marker of both 3SG and 3PL for a very long time. It was used as marker of 3SG only in the extinct Germanic language Gothic, in the 6th century, but came into use as a marker of both 3SG and 3PL shortly thereafter, in the 7th or 8th century (van Bree, 1987).
Therefore, in our study, we controlled for properties of the coda: Half of the verbs ended in a sonorant segment (i.e., nasal or vowel), the other half ended in a non-sonorant segment (i.e., obstruent or fricative). None of the verbs had consonant clusters in the coda of the verb stem, to avoid phonological problems with consonant clusters in the 3SG condition. All verbs had monosyllabic stems and were regular, high-frequency verbs that were selected from two parental vocabulary checklists: the Dutch version of the MacArthur Communicative Development Inventory (Zink & Lejaegere, 2002) and Lexilijst Nederlands (Schlichting & Spelberg, 2002). Moreover, they all had low and very comparable age of acquisition ratings according to an age of acquisition frequency list based on 30,000 Dutch words (M = 5.03, SD = 0.74, range = 3.34 – 5.78) (Brysbaert et al., in press). The filler items tested children’s understanding of lexical contrasts, that is, the meaning of prepositions (e.g., *Ze staat voor/achter het hek* ‘She is standing in front of/behind the fence) or verbs (e.g., *Ze staat/zit op de kist* ‘She is standing/sitting on the box). All items were approximately equal in length, containing four or five words and five or six syllables. For a list of the items, see the Appendix.

The task started with three practice trials. In the first practice trial, children were presented with a lexical contrast, to familiarize them with the procedure. They then received two practice trials presenting pairs of one-actor and two-actor pictures. The aim of these trials was to draw children’s attention to the crucial contrast between the two pictures. In order not to provide them with the targeted structures, children were presented with simple, verbless sentences such as ‘one girl on a seesaw’ with as the accommodating pictures a picture of one girl sitting on a seesaw and a picture of two girls sitting on a seesaw. If children selected the incorrect picture in these practice trials, they were given feedback such that they noticed and understood that the only difference between the two pictures was a difference in the number of actors performing an action.
The items were presented in pseudo-randomized lists in which no more than two items of the same type followed one another. Lists were counterbalanced across participants, such that no child heard the 3SG and 3PL version of an item. Also, to control for left/right response bias, additional lists were created in which the presentation of the pictures (left vs. right) was varied. Forward and backward versions of all lists were used to control for any effects of presentation order such as fatigue.

**Production task.** Production of subject-verb agreement was assessed with an elicited production task in which children were asked to complete sentences produced by the experimenter. In this task, children were presented with pairs of two pictures that presented the same singular/plural contrast as in the comprehension task. To elicit agreement from the children, they were provided with a prompt sentence and asked to finish this sentence. Specifically, the experimenter described what happened on the first picture of a pair and then prompted children to describe the second picture. The following is an example of a prompt sentence: ‘Look! These girls stroke the horse [points to first picture in Figure 2] and this girl ...? [points to second picture in Figure 2].’

[Insert Figure 2 about here]

Note that the second picture in this pair is the same as the item used in the comprehension task. This was done to minimize differences between the two tasks in terms of stimulus materials, both in terms of the verbs used and the picture materials. To make sure that children’s performance in the sentence completion task was not affected by the

---

4 The demonstrative *dit* is used in Dutch with singular neuter nouns, whereas *deze* is used with plural nouns (regardless of gender) and singular common nouns. Hence, the demonstrative pronoun could provide the child with a cue as to the singularity/plurality of the subject. However, due to the Dutch system, there was no way to avoid having some marking of number on the subject DP. For instance, in the case of a definite determiner, *meisje* ‘girl’ is preceded by *het* ‘the’ that marks singular, neuter nouns, while *meisjes* ‘girls’ is preceded by *de* ‘the’ that marks plural nouns (regardless of gender). In the case of an indefinite determiner, *meisje* ‘girl’ is preceded by *een* ‘a’ that marks singularity while the plural indefinite equivalent would be the bare noun *meisjes* ‘girls’.
For Peer Review

Running head: ASYMMETRIES IN SUBJECT-VERB AGREEMENT

comprehension task, the items were counterbalanced across the two tasks such that children did not receive the exact same item in both tasks. Specifically, if they had been presented with a 3SG sentence of a given item in the comprehension task, they were presented with the 3PL sentence of that item in the production task.

As illustrated in Figure 2, the pictures depict an action being performed on different objects (i.e., a horse vs. a dog). The reason for having a contrast in objects between the two pictures of a pair was twofold. First, we expected the children to use sentences containing verbs and objects instead of elliptic utterances (e.g., ‘This girl too’) if they had to describe a contrast between objects. Second, elicitation of an object was considered important given that Dutch children are known to go through a stage in which they produce infinitival clauses (‘root infinitives’) instead of full-fledged finite sentences, such as Mama koekje eten ‘Mommy cookie eat’ rather than Mama eet [een] koekje ‘Mommy eats a cookie’ (Jordens, 1990; Wijnen & Verrips, 1998). Crucially, use of a root infinitive is not considered to be an agreement error (Poeppel & Wexler, 1993). In Dutch, the infinitive is homophonous with the plural as it ends in ‘-en’. One way to distinguish between finite forms and root infinitives is to determine a verb’s position relative to the object: whereas finite forms precede the object, infinitives follow the object in Dutch. So, eliciting objects from the children was considered important to be able to tell apart plural and infinitival verbs.

The production task contained the same twelve experimental items and six filler items as the comprehension task. It started with two practice trials to familiarize children with the procedure. Experimental items were presented in counterbalanced lists in which no more than two items of the same type were presented after one another. Of each list, forward and backward lists were created to avoid order effects.

Procedure
Children were tested individually by trained assistants in a quiet room at their day care centers or at home. In both tasks, assistants provided children with one more trial if children did not respond to the first one. If children did not respond to the second trial either, the item was scored as ‘no response’. The comprehension task always preceded the production task. Although this may have biased children’s performance in favor of production, we decided not to counterbalance the order of the tasks, but always start with the less challenging comprehension task. Since, in the production task, a more active role is asked from the child, we feared that starting with this task would make the two- and three-year-old children shy or unwilling to participate. To reduce any transfer effects from the comprehension to the production task, two very different tasks (i.e., the Dutch Peabody Picture Vocabulary Test and a nonword repetition test) intervened between the two tasks. In addition, the items of the comprehension and production task were counterbalanced to make sure that children were not presented with the exact same items, as explained above. The entire session lasted approximately 30 minutes. At the end of the session, children received a small gift.

Coding and Analyses

Children’s responses in the comprehension task were scored online. Responses could be correct, incorrect, or null responses (2% of all responses). The responses in the production task were transcribed on the basis of video recordings, and coded for (i) verb type (target, semantically related, semantically unrelated, auxiliary, none), (ii) whether there was an object or not, (iii) the order in which the verb and the object appeared (target-like, or non-target-like, e.g., root infinitives), and (iv), agreement marking on the verb (‘-t’, ‘-en’, null, other, e.g., past participles). Scores were calculated as the percentage of correct responses out of the total number of correct and incorrect responses for each child. For production, scores were based on utterances that contained a target or semantically related verb and an object in target-like order only, that is, with the verb in V2 position.
Repeated-measures ANOVAs were run on children’s scores on the comprehension and production tasks to test for effects of condition (singular vs. plural) and group (two- vs. three-year-olds). Also, one sample t-tests with 50% correct as the reference level were run on the data of the comprehension task to see whether scores were significantly above chance.

A large number of responses in the production task could not be analyzed for subject-verb agreement, as they did not contain a verb at all or did not contain a verb in V2 position. To avoid calculating percentages correct on the basis of few responses, only children who produced at least three analyzable (V2 utterances) per condition were taken into account in the ANOVA. As this reduced sample size (and hence the power of the ANOVA), in a second step, an additional logistic mixed-effect regression analysis was run. In this analysis, group and condition were entered as predictor variables and children’s production accuracy on each item was entered as the (categorical) dependent variable (i.e., incorrect vs. correct). Subjects and items were entered as random factors. The aim of this analysis was to see if differences between groups and/or conditions were significant in an analysis on the basis of all data, rather than a small subset of the data, as logistic mixed-effect regression can take into account multiple responses of participants, including those with few responses.

Finally, to explore the relationship between comprehension and production, a Pearson correlation between children’s comprehension and production scores was calculated. Subsequently, individual children’s scores on both tasks were inspected to see how many children performed better on production than on comprehension, how many children performed vice versa, and how many performed alike on both tasks.

**Results**

**Comprehension**

Mean accuracy scores (in percentages correct) and standard deviations on the comprehension task are presented in Table 2 for the two age groups separately.
A repeated-measures ANOVA with condition (singular vs. plural) as the within-subjects factor and group (two- vs. three-year-olds) as the between-subjects factor showed a main effect of condition \( (F(1,68) = 15.19, p < .001, \eta^2_p = .18) \) and a main effect of group \( (F(1,68) = 5.98, p < .05, \eta^2_p = .08) \). No interaction effect was found \( (p > .1) \). These results indicate that both groups performed better on the plural than singular, with the three-year-olds outperforming the two-year-olds in both conditions. A t-test comparing performance against the 50%-chance level showed that, in the singular condition, performance was not above chance in either of the two groups (\( t(1,28) = -1.44, p > .05, \) Cohen’s \( d = 0.54 \) for the two-year-olds; \( t(1,40) = .11, p > .1, \) Cohen’s \( d = 0.03 \) for the three-year-olds). In the plural condition, however, both groups performed significantly above chance, with the effect being larger for the three-year-olds (\( t(1,40) = 5.34, p < .001, \) Cohen’s \( d = 1.67 \)) than for the two-year-olds (\( t(1,28) = 2.39, p < .05, \) Cohen’s \( d = 0.90 \)).

**Production**

An overview of the types of responses given in the elicited production task is presented in Table 3.

In both groups, most responses contained an object and a verb in target-like order (i.e., verb in V2 position) that could be analyzed for subject-verb agreement (69.1% for the two-year-olds; 89.5% for the three-year-olds). In the youngest age group, incomplete responses without a verb or with a verb in non-target position were also relatively frequent.
To avoid calculating accuracy scores for children on the basis of few data points, scores were calculated for children with at least three analyzable responses per condition. Accuracy scores (in percentages correct) are presented in Table 4 for the two groups separately.

A repeated-measures ANOVA with condition as within-subjects factor and group as between-subjects factor showed no effects of condition or group and no interactions (all $p$s > .1). Effect sizes were $\eta^2_p = .03$ for both factors, however, suggesting that the lack of significant results may be due to a lack of power, as $p$-values are sensitive to sample size, but effect size is not. As can be seen in Table 4, there was a fairly small sample of two-year-olds, which may have led to insufficient power in the analysis.

To examine whether the above pattern of results would hold if more children were taken into account, an additional analysis was done in which responses from all children who produced at least one response with a verb in V2 position were taken into account, rather than only those responses from children who produced a minimal number of three analyzable utterances per condition. Specifically, we performed a linear logistic mixed-effect regression analysis with children’s accuracy on each item as the dependent variable ($0 =$ incorrect, $1 =$ correct), and group and condition as the predictor variables. Items and participants were entered as random factors in this analysis (see Method section). Mean accuracy scores for the two groups based on this larger sample of children are shown in Table 5. Table 6 presents the results of the logistic regression analysis.
Both group and condition have marginally significant effects on children’s production accuracy, such that there is a trend for (i) the three-year-olds to outperform the two-year-olds, and (ii) performance to be better on the plural than on the singular. These results match the data in Table 4 above for the smaller group that showed numerically (but not statistically) higher accuracy for the three- than two-year-olds and for the plural than for the singular.

Overall, the production data in Table 5 show much lower accuracy than in the elicited production study on Dutch-speaking three-year-olds by Polišenská (2010) (see Introduction). One possible reason for this discrepancy may be a difference in design across studies. Whereas Polišenská did not provide children with a target verb in the prompt sentence, the current participants were provided with a verb form by the experimenter and then had to produce the same verb but in a different form in their own responses. Possibly, this task of producing the same verb as the experimenter but in a different form made the task more challenging for the children. If so, one would expect the current participants to often repeat the verb form produced by the experimenter, rather than leave out the inflection morpheme, a common agreement error in Dutch children (Blom, 2003; Blom, Polišenská, & Weerman, 2006).

The data in Table 7 suggest that this is indeed what happened. This table shows the frequency of the types of agreement errors made in both conditions for all children who produced at least one error involving an incorrect agreement morpheme (20 two-year-olds and 31 three-year-olds). In the singular condition, agreement errors involved verbs ending in ‘-en’ more often than verbs ending in a 0-morpheme in both the two-year-olds (38% vs. 62%) and three-year-olds (40% vs. 60%). In the plural condition, agreement errors involved verb forms
ending in ‘-t’ more often than forms ending in -0, both in the two-year-olds (27% vs. 73%) and three-year-olds (16% vs. 84%).

[Insert Table 7 about here]

**Relation between comprehension and production**

To investigate how children’s performance on the two tasks was interrelated, we correlated children’s accuracy scores on both tasks. As in the previous analyses, production scores were based on children’s agreement marking on analyzable structures only, and children had to produce at least three V2 structures per condition in order to be included in the analysis. A moderate, positive significant relationship between comprehension and production was found ($r(52) = .38$, $p < .01$). An anonymous reviewer pointed out that this analysis may be problematic as children did not produce analyzable utterances for all verbs in the production task, hence different verbs may be compared across tasks. Therefore, we repeated our analysis for the children who responded to at least ten items in the production task (as a stricter criterion of all twelve utterances would result in a too small sample of only six children). The correlation for this sample of children with (nearly) complete overlap in verbs responded to across tasks, yielded a positive and significant correlation between comprehension and production that was slightly larger than found above ($r(31) = .43$, $p = .01$).

Closer inspection of the data showed that of all 52 children, 30 children performed at least 10% less accurately on the comprehension task than on the production task (58%), 11 children scored more or less the same on both tasks (i.e., less than 10% difference in scores) (21%), and 11 children performed at least 10% more accurately on the comprehension task than on the production task (21%). No differences in distribution were found between the two age groups.
In summary, the present results showed that Dutch two- and three-year old children performed significantly above chance on 3PL, but not on 3SG, in a picture selection task. Elicited production data from the same children showed that both age groups obtained accuracy scores between 64% and 78% on 3SG and 3PL. Analyses on the full set of analyzable responses showed, moreover, that there was a trend for both groups to perform better on the plural than on the singular. Accuracy in production was relatively low, however, as compared with earlier studies. This may be due to children repeating the verb form provided by the experimenter (as a prime) relatively often, as suggested by a comparison of error rates: in both conditions, children used the form produced in the prompt sentence more often than a null form, not produced by the experimenter. Finally, a moderate, significant correlation ($r = .38$) between children’s comprehension and production performance was found. Closer examination of the individual data showed that over half of the children (58%) scored better on comprehension than on production, whereas the remaining children scored better on production, or performed more or less alike on both tasks.

**Discussion**

The results of this study indicate that the comprehension of verbal number agreement in Dutch emerges at two years, but follows an asymmetrical pattern, as children this young show comprehension of the 3PL, but not the 3SG morpheme. At three years, performance is significantly higher, but shows the same pattern, suggesting that the comprehension of 3SG develops more slowly. These findings differ from those on English-, Spanish-, German- and Xhosa-speaking children who do not show above-chance performance on either 3SG or 3PL before the age of five in very similar comprehension tasks. They also differ from findings for French showing that, as early as at 2;6 years, French-speaking children show comprehension of a special case of number agreement (liaison) of both 3SG and 3PL. So, in terms of the rate of acquisition of verbal number agreement, Dutch is in between languages such as Spanish,
English, and German that mark the singular and/or plural with monosyllabic suffixes such as /s/, /n/ or /t/, and French that incidentally marks plural with the special liaison morpheme /z/.

The finding that Dutch children’s comprehension of ‘-en’ (but not ‘-t’) emerges as early as at two years supports the idea that morphemes that are perceptually salient and have a high cue reliability are comprehended earlier than morphemes that are less salient and provide less reliable cues (Brown, 1973; Legendre et al., 2013; MacWhinney et al., 1984). On the basis of these data, we cannot tell apart, however, which one of these factors is more important given that Dutch ‘-en’ is both perceptually salient and provides a more reliable cue as compared to ‘-t’. Another factor that may facilitate the acquisition of 3PL as compared to 3SG is that it involves syllabic rather than consonantal inflection, adding an extra syllable to the verb stem. Previous studies have shown that syllabic morphemes are acquired before non-syllabic ones by children with SLI (Leonard & Bertolini, 1998). Also, input frequency of both specific morphemes and verb forms may play a role. In the current study, frequency was not manipulated as a factor and thus could not be investigated. Future research could examine how factors such as phonological salience, cue reliability, syllabicity, and input frequency (type and token frequency) relate to, and possibly interact, in children’s comprehension and production of agreement.

One of the goals of the current study was to see whether there would be an asymmetry between comprehension and production when the methodology to examine comprehension and production was kept as constant as possible. Our results show that such an asymmetry still exists, with production being more advanced than comprehension, but crucially, that it is much smaller than assumed in previous research. The relatively low production rates in the current study as compared to spontaneous speech are in line with results for English that showing English-speaking children reach 90% accuracy in elicited production only by age four (Rice & Wexler, 2002), about one to two years later than in spontaneous speech (Brown,
1973). They also align with previous findings for German showing that 3SG and 3PL production rates, assessed through parental report, were much lower than reported for spontaneous speech (Brandt-Kobele & Höhle, 2010). This suggests that specific task demands play an important role, an observation that was further strengthened by the discrepancy between the current results and those of Polišenská (2010) who found much higher accuracy rates for Dutch three-year-olds, a difference that seems at least in part due to the prompt sentences used.

A drawback of studies looking into the relation between comprehension and production is that direct comparisons between the two domains are difficult, even if, as in the current study, the exact same materials are used to assess both skills. It seems that strong conclusions about a possible gap are only warranted if performance is either very low or very high, such that there is a clear contrast. In the current study, there was a slight above-chance performance in comprehension coupled with 60-70% accuracy in production. What do results like these tell us about a possible comprehension-production asymmetry? This is a difficult question that may lack a clear answer. Yet, we believe that the current results can contribute to a better understanding of the relation between production and comprehension of subject-verb agreement in Dutch in three ways. First, both in comprehension and production, performance on 3PL was more advanced than on 3SG (either statistically or numerically), which suggests that perceptual salience and/or cue reliability are important for both domains. Second, comprehension and production were only moderately correlated, suggesting parallel development in both domains, but only to some degree. Finally, we found that just over half of the children (58%) obtained higher scores on comprehension than on production, but the other half did not, suggesting considerable individual variation in children’s relative performance on both tasks. Taken together, these results show that there are parallels between comprehension and production of subject-verb agreement in Dutch, with production being
slightly more advanced than comprehension at the group level (but not necessarily so at the individual level), and less clearly so than assumed in previous studies on other languages.

Comprehension-production asymmetries have also been found for other areas of language acquisition, such as the production and comprehension of the German particle *auch* ‘also’ (Höhle, et al., 2009; Hüttner, et al., 2004; Müller et al., 2009) and pronouns in English, Dutch, French, and German (Hendriks & Spenader, 2006 for English; Weissenborn, Kail, & Friederici, 1990 for Dutch, French, and German). This raises important issues about the origin of such asymmetries. Previous accounts have proposed that extra-linguistic factors play a role such as limited cognitive factors (Grodzinsky & Reinhart, 1993) or pragmatic factors (Thornton & Wexler, 1999), or explained such asymmetries within an Optimality Theory framework (Hendriks, 2008). Others have argued that the asymmetries are due to methodology, such as different demands of comprehension and production tasks (Hirsh-Pasek & Golinkoff, 1996) or side-effects of the comprehension tasks used (Brandt-Kobele & Höhle, 2010). The present results suggest that asymmetries are still found if task materials across the two domains are kept as similar as possible and if comprehension and production are assessed in the same children. However the results also show that there are clear individual differences. In the current study, about half of the children performed better on comprehension than on production (58%), but the remaining children did not. Also, even though the use of the materials in both domains provides a relatively good test of comprehension-production asymmetries, it does not preclude asymmetrical findings due to diverging methodological demands: listening to a sentence and choosing one out of two minimally differing pictures is likely to call on different processes than producing a sentence containing a correct agreement morpheme. One explanation for the current results that does take into account individual variation, then, is that picture selection tasks place high demands on children’s information processing system, and that children vary in their information processing skills. In order to
perform this task, children have to listen to a sentence, keep this sentence in mind, and – at the same time – visually process two pictures (as well as the difference between these), and then make a motor response. A question left for further research is whether a comprehension-production asymmetry would still be found if a less demanding comprehension task such as, for example, a preferential looking task, was paired with a relatively demanding production task such as the current one.

Conversely, one factor that could contribute to the late comprehension of verbal number agreement – next to or instead of task demands – is its redundant character. Singularity/plurality is typically also encoded on the subject in non-pro-drop languages such as Dutch, so children may be able to convey a singular or plural meaning without fully grasping the meaning expressed by agreement morphemes (provided they know how to use plural morphology on nouns). On this account, the inflected verb forms they produce would be taken over directly from the input without being fully analyzed, as part of larger structures (Theakston, Lieven, & Tomasello, 2003), and be comprehended only later in development. Studies using nonsense verbs (Polišenská, 2010) or detailed analyses on productivity of morphemes (Gathercole et al., 2002) suggest, however, that subject-verb agreement marking may be used productively from early on, at least in some children, which is hard to reconcile with this idea.

Before we conclude that Dutch children’s comprehension of the 3PL morpheme emerges as young as at age two, it is worth considering two alternative interpretations of our results. According to the first interpretation, children would show above-chance performance on 3PL but not 3SG because of an overall preference for 3PL pictures, regardless of condition. Recall from above that Brandt-Kobele and Höhle (2010) found that German three-to four-year-olds had a preference for plural pictures in a preferential looking experiment containing the same type of sentences as studied here, also in the baseline phase of the
experiment (i.e., prior to the presentation of the sentences). To explain this preference, they propose that the two-actor pictures attract more attention from young children because they are informationally more complex. The current results are compatible with such an overall preference, as children selected the plural pictures more often than the singular pictures in both conditions. However, it does not seem likely that our results are due to an overall preference for two-actor pictures for two reasons. First, Brandt-Kobele and Höhle found that the preference for two-actor pictures was only visible in eye movements and not in children’s pointing in a picture selection task. This makes it unlikely that Dutch children would show such a preference in pointing. Second, it is not clear why such a tendency would be stronger for three-year-olds than for two-year-olds, as found in the current study, because children’s information processing skills should increase rather than decrease with age. For these reasons, we think it is safe to assume that our comprehension data reflect Dutch children’s ability to comprehend verbal number agreement rather than a non-linguistic overall preference for informationally more complex pictures.

The second, alternative interpretation of our comprehension results would hold that children opted for the plural pictures most because a singular sentence is also compatible with two-actor pictures (i.e., for one out of the two actors depicted), but not vice versa. While we cannot rule out this possibility on the basis of the present data, it does not seem likely either. First, French-speaking could match 3SG as well as 3PL-sentences to their correct pictures in a similar pointing task involving a contrast between one and two actors already at 30 months of age (Legendre et al., 2010). Second, in a preferential looking experiment, both French- and German-speaking children looked longer to the matching pictures in both the 3SG and 3PL conditions, revealing sensitivity to 3PL as well as 3SG (Legendre et al., 2010; Brandt-Kobele & Höhle, 2010). These results show that young children may associate 3SG-sentences with
one-actor pictures, making it unlikely that Dutch children of an older age would not be able to do so.

A question left for future research is when Dutch children will show full mastery of verbal number agreement in comprehension and production. For comprehension, closer inspection of our data showed that, in our sample, even children who were nearly four years old often did not show above-chance comprehension of 3SG. In our total sample, only one child showed 100% mastery on 3SG and 3PL in the comprehension task, and three children showed near-mastery (over 90% correct on both conditions). For production, ten children showed 100% mastery on 3SG and 3PL in the production task, and two children showed 90% mastery on this task, all of them being three-year-olds. In the two-year-olds’, responses often did not involve the targeted (verb-object) constructions, but lacked a verb or an object or had an incorrect order (i.e., root infinitives). While this posed a problem for our analyses in terms of power, the two-year-olds’ frequent omission of (finite) verbs may be interesting in itself as it could reveal avoidance of subject-verb agreement and reflect, at least in part, young children’s difficulties with producing subject-verb agreement. Future research could investigate whether comprehension of the 3SG morpheme is earlier, later or follows the same time course as in previously investigated languages such as English and German in which morphemes are also phonologically not very salient and the same morpheme can be used to mark several functions within the paradigm. Ideally, this research involves older children, at least up until age five, to be able to track the acquisition process from beginning to end and compare the results cross-linguistically across age groups.

Also, future research could use more advanced methods of studying comprehension than the current study. As outlined above, earlier studies that coupled (manual) forced-choice picture selection tasks with preferential looking tasks (Brandt-Kobele & Höhle, 2010; Legendre et al., 2010) have shown that the latter type of tasks may reveal knowledge of
subject-verb agreement in young children that does not become apparent in children’s pointing behavior. In the current study, we opted for a picture selection task, because our aim was to extend earlier studies on English, German, and French that all used picture selection tasks to a new language. There are, however, a few drawbacks to using picture selection as a measure of comprehension. One is that performance is compared against chance, and one may wonder whether above-chance performance is a good indicator of comprehension. Another limitation of the task is its demanding character: Children not only have to integrate auditory and visual information but also have to make an explicit judgment about sentences in the form of a motor response. This demanding task, together with its limited informativeness (i.e., comparing against chance level as a measure of comprehension), may make picture selection less than ideal for studying comprehension in young children. Future research on comprehension-production asymmetries should explore the use of other methods to assess comprehension, such as eye tracking or neurocognitive methods, to obtain data that reflect children’s comprehension abilities more closely.

Taken together, the results of this study show that Dutch children aged as young as two and three years show above-chance comprehension of the agreement morpheme marking 3PL, but not 3SG, supporting previous claims that phonological salience and cue reliability foster comprehension. As suggested in previous work, there was an asymmetry between comprehension and production, despite complete overlap in task stimuli across the two domains. This asymmetry, however, was much smaller than assumed in previous studies on the basis of spontaneous speech. This suggests that, apart from linguistic properties of the target language, specific task demands have a clear impact on children’s performance.
References


Weissenborn, J., Kail, M., & Friederici, A. (1990) Language particular or language
independent factors in acquisition? Children’s comprehension of object pronouns in Dutch, French and German. *First Language, 10*, 141-166.


Appendix

Experimental items of the comprehension and production tasks:

1. Ze aait/aaien het hondje  ‘She/they strokes/stroke the dog’
2. Ze bouwt/bouwen de toren  ‘She/they builds/build the tower’
3. Ze duwt/duwen de kar  ‘She/they pushes/push the cart’
4. Ze gooit/gooien de bal  ‘She/they throws/throw the ball’
5. Ze kamt/kammen de pop  ‘She/they combs/comb the doll’
6. Ze klimt/klimmen op de trap  ‘She/they climbs/climb the stair’
7. Ze kijkt/kijken naar de TV  ‘She/they watches/watch television’
8. Ze loopt/lopen op de straat  ‘She/they walks/walk in the street’
9. Ze slaapt/slapen in het bed  ‘She/they sleeps/sleep in the bed’
10. Ze zoekt/zoeken de fiets  ‘She/they looks for/look for the bike’
11. Ze draagt/dragen de mand  ‘She/they carries/carry the basket’
12. Ze leest/lezen het boek  ‘She/they reads/read the book’
Table 1

Agreement Paradigm in Dutch (Present)

<table>
<thead>
<tr>
<th></th>
<th>Dutch example</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>Ik zing-ø</td>
<td>I sing</td>
</tr>
<tr>
<td>2SG</td>
<td>Jij zing-t</td>
<td>You sing</td>
</tr>
<tr>
<td>3SG</td>
<td>Hij/zij zing-t</td>
<td>He/she sings</td>
</tr>
<tr>
<td>1PL</td>
<td>Wij zing-en</td>
<td>We sing</td>
</tr>
<tr>
<td>2PL</td>
<td>Jullie zing-en</td>
<td>You sing</td>
</tr>
<tr>
<td>3PL</td>
<td>Zij zing-en</td>
<td>They sing</td>
</tr>
</tbody>
</table>
Table 2

*Mean Scores and Standard Deviations on the Comprehension Task for the Two- and Three-Year-Olds (in %)*

<table>
<thead>
<tr>
<th></th>
<th>Two-year-olds</th>
<th>Three-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Singular</td>
<td>44.3</td>
<td>21.5</td>
</tr>
<tr>
<td>Plural</td>
<td>58.1</td>
<td>18.2</td>
</tr>
</tbody>
</table>
Table 3

Response Types in the Production Task (in % and (N))

<table>
<thead>
<tr>
<th></th>
<th>Two-year-olds (N = 29)</th>
<th>Three-year-olds (N = 41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical verb, no object</td>
<td>12.7 (30)</td>
<td>2.4 (10)</td>
</tr>
<tr>
<td>Lexical verb, object (non-target order)</td>
<td>18.2 (43)</td>
<td>8.1 (34)</td>
</tr>
<tr>
<td>Lexical verb, object (target order)</td>
<td>69.1 (163)</td>
<td>89.5 (374)</td>
</tr>
<tr>
<td>Auxiliary verb</td>
<td>2.6 (9)</td>
<td>6.1 (30)</td>
</tr>
<tr>
<td>Other (no verb, unintelligible)</td>
<td>19.0 (66)</td>
<td>6.1 (30)</td>
</tr>
<tr>
<td>No response</td>
<td>10.6 (37)</td>
<td>2.8 (14)</td>
</tr>
<tr>
<td>Total number of responses</td>
<td>348</td>
<td>492</td>
</tr>
</tbody>
</table>
Table 4

*Mean Scores and Standard Deviations on the Production Task for the Two- and Three-Year-Olds (in %)*

<table>
<thead>
<tr>
<th></th>
<th>Two-year-olds</th>
<th>Three-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Singular</td>
<td>64.2</td>
<td>27.1</td>
</tr>
<tr>
<td>Plural</td>
<td>71.2</td>
<td>26.1</td>
</tr>
</tbody>
</table>
Table 5

Correct and Incorrect Verb Forms in the Production Task per Condition for the Two- and Three-Year-Olds (in % and (N))

<table>
<thead>
<tr>
<th></th>
<th>Two-year-olds (N = 25)</th>
<th>Three-year-olds (N = 41)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Singular</td>
<td>55.3 (42)</td>
<td>44.7 (34)</td>
</tr>
<tr>
<td>Plural</td>
<td>66.3 (59)</td>
<td>33.7 (30)</td>
</tr>
</tbody>
</table>
Table 6

Model Estimates, Standard Error (SE), Z- and P-values of Mixed-Effect Regression

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficient</th>
<th>SE</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.65</td>
<td>0.32</td>
<td>2.02</td>
<td>.044</td>
</tr>
<tr>
<td>Group</td>
<td>0.71</td>
<td>0.37</td>
<td>1.90</td>
<td>.057</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.42</td>
<td>0.22</td>
<td>-1.89</td>
<td>.058</td>
</tr>
</tbody>
</table>
Table 7

*Error Types per Condition in Production Task (in % and (N))*

<table>
<thead>
<tr>
<th></th>
<th>Two-year-olds (N = 20)</th>
<th>Three-year-olds (N = 31)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0 (Sg)</td>
<td>- (Pl)</td>
</tr>
<tr>
<td>Singular</td>
<td>38.2 (13)</td>
<td>61.8 (21)</td>
</tr>
<tr>
<td>Plural</td>
<td>26.7 (8)</td>
<td>73.3 (22)</td>
</tr>
<tr>
<td></td>
<td>61.8 (21)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>39.7 (23)</td>
<td>15.6 (7)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>84.4 (38)</td>
</tr>
<tr>
<td></td>
<td>60.3 (35)</td>
<td>-</td>
</tr>
</tbody>
</table>
Singular condition:
Ze aait het hondje (‘She strokes the dog’)

Plural condition:
Ze aaïen het hondje (‘They stroke the dog’)

Figure 1.
Example item of the picture selection task
Singular condition:  

Plural condition:  

Figure 2.

Example item of the sentence-completion task