

Assessing the role of current and CUMULATIVE exposure in simultaneous bilingual acquisition: The case of Dutch gender*

SHARON UNSWORTH
Utrecht University

(Received: March 16, 2011; final revision received: May 4, 2012; accepted: May 16, 2012; first published online 6 September 2012)

This paper investigates the role of amount of current and CUMULATIVE exposure in bilingual development and ultimate attainment by exploring the extent to which simultaneous bilingual children's knowledge of grammatical gender is affected by current and previous amount of exposure, including in the early years. Elicited production and grammaticality judgement data collected from 136 English–Dutch-speaking bilingual children aged between three and 17 years are used to examine the lexical and grammatical aspects of Dutch gender, viz. definite determiners and adjectival inflection. It is argued that the results are more consistent with a rule-based than a piecemeal approach to acquisition (Blom, Polišenská & Weerman, 2008a; Gathercole & Thomas, 2005, 2009), and that non-target performance on the production task can be explained by the Missing Surface Inflection Hypothesis (Haznedar & Schwartz, 1997; Prévost & White, 2000; Weerman, Duijnmeijer & Orgassa, 2011).

Keywords: simultaneous bilingual acquisition, exposure effects, Dutch, grammatical gender, cumulative length of exposure

1. Introduction

One of the sources of variation in bilingual populations is the amount and type of language to which children are exposed. Bilingual children, by definition, have to divide their waking hours between their two languages, and consequently, are probably almost always exposed to less input than monolinguals (Paradis & Genesee, 1996). While a number of studies have shown that amount of exposure is indeed a significant predictor of certain language outcomes in bilingual children, there is little consensus about which linguistic domains should be affected or to what extent (see e.g., Sorace, 2011, for suggestions). Furthermore, research into the effects of amount of exposure typically focuses on the child's current

situation; exposure accumulated over time has received little attention, and discussion of the role of (amount of) exposure in the early childhood years is typically confined to successive bilinguals.

This paper examines the effects of amount of exposure on the acquisition of grammatical gender in Dutch in 136 simultaneous bilingual English–Dutch children using elicited production and grammaticality judgement data. As part of this investigation, we assess children's exposure at the current time and introduce the notion of CUMULATIVE length of exposure, a measure intended to capture the sum of bilingual children's language exposure over time and to facilitate more accurate comparisons between bilingual and monolingual language development.

Section 2 reviews previous literature on exposure effects in bilingual acquisition, before we turn to grammatical gender in Dutch in Section 3. Section 4 outlines the research questions and predictions of the current study. The details of how amount of exposure is estimated, how data on grammatical gender are elicited, as well as information concerning participants are all presented in Section 5, before presenting the results in Section 6. Finally, in Sections 7 and 8, we return to the research questions and a more general discussion of the issues most relevant to the present study.

* This research was supported by the Netherlands Organisation for Scientific Research with a VENI Innovational Research Incentives Scheme award to the author and an international programme award to Leonie Cornips. I wish to thank the participants and research assistants, as well as Leonie Cornips, Aafke Hulk, Antonella Sorace and Ianthi Tsimpli for discussion of some of the issues in this paper, Enlli Môn Thomas for comments on an earlier version of this paper, and Harvard Language and Cognition group for feedback on a presentation version. I also thank three anonymous reviewers for their critical and constructive comments.

Address for correspondence:
Department of Modern Languages/Utrecht Institute of Linguistics, Trans 10, 3512JK Utrecht, The Netherlands
s.unsworth@uu.nl

2. Effects of amount of exposure on bilingual acquisition

Various studies have shown that bilingual children's acquisition of vocabulary is affected by amount of exposure (e.g., Pearson, Fernández, Lewedeg & Oller, 1997). A number of studies have also examined the effect of differential amounts of exposure on bilingual children's morphosyntactic development, for various target language properties, including grammatical gender (e.g., Gathercole, 2002a), verbal morphology (e.g., Paradis, 2010a), finiteness (Blom, 2010), mass/count nouns (Gathercole, 2002b), *that*-trace effects (Gathercole, 2002c) and *wh*-questions, passives and definite/indefinite articles (Chondrogianni & Marinis, 2011), and in many cases, amount of exposure has been found to affect rate of acquisition. Thus, differences in amount of input have been shown to affect both bilingual children's language abilities and the rate at which they acquire various linguistic phenomena relative to monolinguals.

Even though exposure effects have been observed, it is worth noting that by and large the morphosyntactic development of simultaneous bilingual children generally patterns similarly to that of monolingual children, both in terms of rate and error types (Genesee & Nicoladis, 2007). In other words, while effects of amount of exposure may be observed, and, in some cases, bilingual children have also been shown to acquire certain properties more slowly or quickly than monolinguals (see e.g., Meisel 2007a), the relationship between amount of language exposure and language development is clearly not a direct one (Gutiérrez-Clellen & Kreiter, 2003). The focus of much of this work is thus to determine what exactly this relationship is, and to what extent this is moderated and/or mediated by other variables.¹

In a large-scale investigation of the linguistic abilities of Spanish–English bilingual children in Miami, the effect of amount of exposure was observed to be greater in the earlier years, i.e., at kindergarten and in grade 2, and by grade 5 this effect was significantly reduced (Oller & Eilers, 2002). Adopting a usage-based approach, Gathercole (2002c) proposes that children need time to reach a “critical mass” in the input, i.e., they need to reach a certain threshold of “exemplars” in order for acquisition to take place. This threshold may vary depending on the transparency and reliability of the input in terms of e.g., form–function mappings. The challenge for such an approach to bilingual acquisition, where amount of exposure is the crucial explanatory factor in the success and timing of the acquisition of certain target language properties, is being able to specify the relative thresholds,

i.e., quantifying the “critical mass” such that specific predictions can be made, a challenge which as yet has not been met (Gathercole & Thomas, 2009, p. 215, fn. 1; see also Ellis, 2006, for relevant discussion concerning the power law of learning).

Nevertheless, the empirical observation that the effects of amount of exposure to the target language appear to diminish over time remains. Gathercole and Thomas (2009, p. 234) furthermore suggest that for the minority language (Welsh) in their study, continual exposure may be needed through the lifespan in order to reach and maintain “nativelike” mastery.

To summarise, there is evidence that certain aspects of bilingual children's linguistic development are affected by the amount of language to which they are exposed, and specific characteristics thereof. The overall goal of this paper is to examine the effect of differential exposure patterns on the acquisition of Dutch gender. In contrast to the acquisition of grammatical gender in many other languages, the acquisition of grammatical gender in Dutch is a long and drawn out process, with monolingual (L1) children making errors until at least age six (e.g., van der Velde, 2003). The following section briefly reviews the relevant properties of the Dutch gender system, and the results of previous studies on its acquisition, focussing in particular on the (potential) role of the amount of exposure in the bilingual context.

3. Acquisition of Dutch gender

3.1 Grammatical gender in Dutch

Dutch has a two-way gender system, distinguishing common from neuter. Grammatical gender is marked on a number of agreeing elements inside and outside the DP, including definite determiners, demonstratives, relative pronouns, first person plural possessives, *wh*-phrases, and attributive adjectives. This is illustrated for definite determiners and attributive adjectives in (1a) for common and (1b) for neuter (see Blom, Polišenská & Unsworth, 2008b, for overview).

- (1) a. de kleine fiets / een kleine fiets / de
 the small bike a small bike the
 kleine fietsen (common)
 small bikes
- b. het kleine huis / een klein huis / de
 the small house a small house the
 kleine huizen (neuter)
 small houses

The gender specification of a given noun in Dutch is generally assumed to be arbitrary (Deutsch & Wijnen, 1985) and although a number of morphosyntactic and semantic cues exist, these are limited and numerous

¹ Although not directly addressed here, bilingual children's language exposure may also vary in type, i.e., in terms of quality as well as quantity.

exceptions exist (Donaldson, 1987; Geerts, Haeseryn, de Rooij & van de Toorn, 1984).² The focus of the present paper is gender-marking on definite determiners and attributive adjectives in indefinite DPs. As shown in (1), common nouns are preceded by the definite determiner *de*, whereas neuter nouns combine with the definite determiner *het*; attributive adjectives are inflected with a schwa in all cases EXCEPT with singular, indefinite neuter nouns.

In order to acquire grammatical gender, children need to know (i) that gender is a grammatical feature instantiated in DPs; (ii) the gender specification of the noun in question, i.e., gender attribution; and (iii) how to mark gender on other elements in the DP, i.e., gender concord or agreement (Carroll, 1989; Meisel, 2009). Following Carstens (2000), it is assumed that all nouns in Dutch are marked with an interpretable gender feature [\pm neuter] which checks (or values) the uninterpretable [μ gender] feature on D and A in either a head–head or a spec–head relation, respectively.³ Subsequently, adopting the distributed morphology approach taken for Dutch by Blom, Poliřenskà and Weerman (2008a), the result of this checking operation is a value which is interpreted by the vocabulary component consisting of lists of partially specified phonological forms (“vocabulary items”) ready to be inserted into the terminal node (Halle & Marantz, 1993). In combination with the Elsewhere Principle (Kiparsky, 1973) and the Subset Principle (Halle, 1997), the lexical insertion rules in (2) and (3), where [\pm attr],[\pm def] and [\pm plur] respectively stand for attributiveness, definiteness and plurality, derive the observed patterns for definite determiners and adjectives in Dutch.

(2) /het/ \leftrightarrow [+def, +neut, –plur]
 /de/ \leftrightarrow [+def]

(3) /Ø/ \leftrightarrow [+attr, –def, +neut, –plur]
 /ə/ \leftrightarrow [+attr]

(Blom et al., 2008a, p. 306, exx. (9) and (7), respectively)

Thus, according to (2), *de* is inserted in all definite contexts, unless the noun is singular and neuter, and similarly, (3) states that the inflected form of the adjective is inserted in all attributive contexts unless the noun is indefinite, singular and neuter. On this analysis, then, the acquisition of gender-marking on definite determiners and adjectives involves acquiring the topmost rules in (2) and (3).

² As an anonymous reviewer points out, it is possible that certain agreeing elements, and in particular determiners, may be privileged cues.

³ See Roodenburg and Hulk (2008) for an alternative approach.

3.2 Previous studies on monolingual/bilingual acquisition of Dutch gender

Previous research on the acquisition of Dutch gender shows that bilingual children produce similar errors to monolingual children, overgeneralising *de* with neuter nouns, producing non-target combinations such as *de_{COMMON} huis_{NEUTER}}*, and overgeneralising the inflected form of the adjective, as in **een kleine huis* (Blom et al., 2008a; Cornips, van der Hoek & Verwer, 2006; De Houwer, 1990). Errors in the other direction occur infrequently, and while monolingual children eventually acquire the target system, it is unclear whether bilingual children ever proceed beyond this stage of overgeneralisation. Note that in several of these studies (Brouwer et al., 2008; Cornips & Hulk, 2006; Cornips et al., 2006), it is also unclear whether the children should be classified as simultaneous or successive bilingual children. Whilst we might assume that simultaneous bilingual children will eventually approximate the same level of ultimate attainment as monolingual children, the nature of gender in Dutch and the importance of input for its acquisition mean that this assumption may be questionable in the present context. There is as yet no study which investigates this issue directly for Dutch; the present study seeks to fill this gap.

It is possible that the lower accuracy rates for bilinguals are the result of (at least some of) the children in previous studies first being exposed to Dutch after birth, i.e., an effect of age of onset. A recent study by Unsworth, Argyri, Cornips, Hulk, Sorace and Tsimpli (in press) suggests however that the relevant factor is amount of exposure rather than age of onset (see also Unsworth, in press). As the authors note, this finding may be due to the limited cues for neuter gender available in the input. These result from the lack of systematic morpho-phonological marking, common nouns outnumbering neuter by approximately 2:1 (Van Berkum, 1996), the common form appearing wherever the gender distinction is neutralised e.g., plurals, and the lexical form *het* serving several other functions, including e.g., as a pronominal form, in impersonal constructions, with nominalised infinitives and with predicative superlatives (Roodenburg & Hulk, 2008). In other words, the grammatical gender system in Dutch may be considered “opaque”, with the consequence that the specification of gender in Dutch must to a certain extent occur on a word-by-word basis (Blom et al., 2008a; Unsworth, 2008; Weerman, Duijnmeijer & Orgassa, 2011).

This finding is furthermore in line with Blom et al. (2008a), who propose that, at least in the early stages of acquisition, monolingual and bilingual/L2 children acquire gender-marking on definite determiners via lexical learning in the form of “lexical frames” induced on the basis of the input. As these authors note,

when children produce a congruent determiner–noun combination, it is impossible to know whether this results from such lexical (item-by-item) learning or from a grammar-based strategy incorporating abstract features and rules such as those discussed above. When congruent (determiner–)adjective–noun combinations are consistently produced, however, it is likely that these are due to grammatical agreement within the DP; clearly, children must additionally acquire the topmost rule in (3), but once “learners activate [\pm neuter], it is . . . expected that this will influence their linguistic performance in all . . . gender domains” (Blom et al., 2008a, p. 308). Interestingly, collating results from a number of studies, Blom et al. (2008a, p. 323) observe that the bilingual/L2 children studied thus far appear not to acquire this rule and they speculate that this may be because its acquisition requires “a lengthy period of substantial exposure [to] compensate for weak statistical properties in the input”, i.e., the reduced amount of input to which these children are exposed (either due to a late(r) start in the case of the L2 children in their study or more generally to the nature of the bilingual situation) means that the relevant “critical mass” of information to deduce this rule cannot be attained within a critical period which may end around age six or seven (Meisel, 2007b; see also Meisel, 2009).

The idea that bilingual children may need sufficient exposure in order to acquire complex or “opaque” properties of the target language is also put forward, albeit from a different approach, by Gathercole and Thomas (2005). In their study (see also Gathercole & Thomas, 2009; Thomas & Gathercole, 2007), these authors observe that bilingual children with the least exposure to Welsh perform poorly on the more “opaque” aspects of the gender system, e.g., where multiple form–function pairings exist and where the application of gender-marking is restricted to certain contexts and nouns (see also Kupisch, Müller & Cantone, 2002). In their conclusion, Gathercole and Thomas (2005) speculate that the acquisition of these aspects of the gender system in Welsh may never take place because for children with comparatively little exposure acquisition may be “timed off the map”, possibly within a critical or sensitive period. While this line of reasoning is similar to that put forward for the acquisition of gender-marking on adjectives by Blom et al. (2008a), it is crucial to note that while the latter authors consider children to ultimately acquire and use rules which employ abstract grammatical features, Gathercole and Thomas (2005, 2009; Thomas & Gathercole 2007) do not; rather, according to these authors, in the acquisition of opaque gender systems, such as Welsh, children adopt a piecemeal, item-by-item approach for all aspects of the system.

Another possible explanation for bilingual/L2 children’s poor performance on Dutch gender-marking – both for adjectives and definite determiners – is that

their errors reflect a problem with PRODUCING gender-marked forms rather than being due to a representational deficit, i.e., a failure to acquire the relevant grammatical features (whether this be due to reduced input, a critical or sensitive period, or both). As posited by the Missing Surface Inflection Hypothesis (MSIH; Haznedar & Schwartz, 1997; Prévost & White, 2000), these features are in place but children experience problems spelling them out in production. What this would mean for Dutch gender, following work on L2 Spanish by White and colleagues (e.g., White, Valenzuela, Kozłowska-MacGregor & Leung, 2004), is that assuming the rules given in (2) and (3) above, children would, in certain (in some sense demanding) contexts, resort to the default, less specified form, which would mean inserting *de* instead of *het* and the inflected form of the adjective instead of the bare form (Blom & Vasic, 2011; Unsworth & Hulk, 2009; Weerman et al., 2011).⁴

Two studies have explicitly examined this question for Dutch. Brouwer, Cornips and Hulk (2008) observe that in a grammaticality judgement task where children were asked to evaluate both congruent and incongruent determiner–noun combinations, 11- to 13-year-old bilingual/L2 children demonstrated some sensitivity to gender-marking on determiners, but nevertheless performed less well than their monolingual peers. In a self-paced listening task, Blom and Vasic (2011) find that 6- to 9-year-old bilingual/L2 children similarly showed sensitivity to mismatches in determiner–noun agreement, but at the same time they made errors in production with the same nouns, but this was for diminutive nouns only and thus the results are only partly in line with the MSIH; however, given that the adult control group also failed to perform as expected for non-derived nouns, the possibility that the lack of an effect for children may be task-related cannot be ruled out.

In short, previous research suggests that in a bilingual context, the amount of language to which children are exposed may affect their acquisition of gender-marking on determiners and adjectives in Dutch. However, this question has thus far only been based on general population characteristics rather than an assessment of the input situation of individual children. In addition, it remains unclear whether bilingual children ever acquire grammatical gender in Dutch, i.e., whether their problems are representational in nature (for whatever reason) or specifically related to production. Furthermore, the relationship between gender-marking on determiners and adjectives has not yet been thoroughly examined in older bilingual children.

⁴ The Dutch gender data illustrate that on the MSIH approach, defaults may involve substitution as well as omission (McCarthy, 2008); as such, the nomenclature MISSING inflection is perhaps misleading.

4. Research questions, hypotheses and predictions

The first research question to be addressed in the present study is the following:

- What is the effect of differential amounts of exposure – now and in the past – on the acquisition of grammatical gender in Dutch by simultaneous bilingual children, and more specifically, are these effects similar for gender-marking on definite determiners and gender-marking on adjectives?

Given previous results, a significant effect of amount of exposure is expected on gender-marking with determiners, specifically with neuter nouns. Furthermore, if amount of exposure is crucial to the acquisition of Dutch gender-marking on determiners, it is expected that when matched for amount of exposure, bilingual children will perform similarly to monolingual children.

With respect to adjectives, the predictions are slightly more complicated. On a rule-based approach, one would expect that any effects of (current and past) amount of exposure on gender agreement be mediated by knowledge of gender attribution, i.e., once children know the appropriate rule (recall (3) above), they will consistently apply it (as observed for monolingual children by Polišenská, 2010). Thus, if – as is common practice in the literature (see Bruhn de Garavito & White, 2002) – we assume that gender-marking on definite determiners can be used as an indicator of gender attribution, and that this is where exposure effects are expected (on any approach), a clear prediction can be made: once knowledge of gender attribution is taken into account, bilingual children's production of gender-marking on adjectives will be less affected by amount of exposure than gender-marking on determiners.

On the piecemeal approach put forward by Gathercole and Thomas (2005, 2009), one would expect to find effects of amount of exposure for accuracy on all aspects of the gender system, i.e., on adjectives as on determiners. Thus, on the assumption that the Dutch gender system is opaque in a similar sense to the gender system in Welsh, it is predicted that for bilingual children, especially those with comparatively limited exposure, acquisition may be “timed off the map”, i.e., they may fail to accrue enough exposure to consistently produce *het* both within and across nouns and to consistently use the uninflected form of the adjective with singular, indefinite neuter nouns. In other words, their ultimate attainment will not be consistent with the target system.

The second research question is as follows:

- What is the source of children's errors in their production of gender-marking on definite determiners and adjectives?

We will explore two possibilities, namely the timing and amount of exposure, and modality (production vs. comprehension).

Blom et al. (2008a, p. 323) speculate that L2 children's failure to acquire the relevant lexical insertion rule for adjectival inflection may be due to reduced input, possibly within a critical period ending at around age six or seven (Meisel, 2007b), which means that children fail to accrue enough evidence to induce this rule. Given that simultaneous bilingual children also have reduced input (compared with monolinguals), it is possible that they too may fail to reach the relevant threshold in the aforementioned timeframe. If this is the case, it is predicted that any failure of older simultaneous bilingual children to demonstrate knowledge of gender-marking on adjectives may be due to insufficient exposure in the early years.

It is also possible that children may produce non-target forms not because they have failed to acquire the relevant grammatical features and rules and/or to specify certain nouns with the target gender feature, but because, following the MSIH (Haznedar & Schwartz, 1997; Prévost & White, 2000), they have a production-specific performance problem. If this is the case, it is predicted that children will be significantly more targetlike on a task which does not involve production, such as a grammaticality judgement task (Blom & Vasic, 2011; Weerman et al., 2011).

5. Method

5.1 Determining amount of exposure

A detailed parental questionnaire (following De Houwer, 2009; Gathercole, p.c.; Gutiérrez-Clellen & Kreiter, 2003; Jia & Aaronson, 2003; Paradis, 2011) was used to estimate children's current amount of exposure, as well as their amount of exposure over time.

Following Gutiérrez-Clellen and Kreiter (2003), amount of exposure was calculated by asking parents to indicate where and with whom the child spent time on an average day in the week and an average day at the weekend, for how long, and which language(s) each person used when addressing the child, using a five-point scale, as well as time spent on extra-curricular activities and the language(s) in which these occurred. Using this information, we made the following calculations: (i) the amount of time each person spends with child multiplied by how much that person speaks Dutch to child, (ii) amount of time child spends at daycare/school multiplied by how much Dutch is spoken at school, (iii) amount of time child spends on extra-curricular activities (namely sports and clubs outside school and after-school care, time spent with friends, watching TV, reading and using the computer (for language-based activities)) multiplied

by how much of these are in Dutch. The total number of hours with language exposure in Dutch is subsequently divided by the child's total number of waking hours to give the overall percentage of current exposure to Dutch per week.

As well as measuring children's current exposure to Dutch, we also examined their exposure over time. As previous literature indicates, and as will become apparent from the results of the above calculations, the amount of exposure varies considerably both among children and within one and the same child over time. As it is identical to chronological age, length of exposure is not usually considered relevant in the study of simultaneous bilingual children. However, given that one year of "bilingual" language exposure is not the same as one year of "monolingual" language exposure, and the amount of exposure varies among bilinguals, any comprehensive evaluation of the role of exposure in this group needs to include an accurate assessment of this variable over time (see Gutiérrez-Clellen & Kreiter, 2003 for – to our knowledge – the only study which has hitherto considered this aspect of bilingual language exposure, albeit measured in a less detailed fashion than in the present study). In the present study, this is achieved using the measure CUMULATIVE length of exposure.

To calculate this measure, the following information was gathered: (i) how much each parent and any other adults living in the home spoke English–Dutch for each one-year period in the child's life so far, using the same scale as for current amount of exposure; and (ii) whether the child attended daycare or school in these periods, and if so, what the language of instruction was there, using the same scale. Using this information, the proportion of each one-year period which included exposure to Dutch was calculated and summed up to give the total amount of exposure to Dutch in years over time.

5.2 Participants

The participants were 136 children raised bilingually in English and Dutch from birth, and aged between three and 17 years at time of testing. They were all resident in the Netherlands at time of testing and the vast majority ($n = 105$) were also born there. All were exposed to both languages at home from birth, usually in a "one parent, one language" situation, although in some families, both parents spoke both languages to the child from birth. There was no history of language delay or impairment.

The current language exposure situation was as follows. For the majority of children ($n = 71$), the mother speaks English most or all of the time and the father Dutch, whereas for 21 children, the pattern is reversed. In 28 families, both parents currently speak English at least 50% of the time, and in 7 families, the same holds for Dutch. There are two one-parent families, one with a

Dutch-speaking mother and one with an English-speaking father. In the remaining 17 families, both parents currently speak both languages more or less equally often when addressing the child.

At the time of testing, most children were attending Dutch-speaking state schools ($n = 93$) or daycare/pre-school ($n = 13$), some were attending an international primary or secondary school where English is the language of instruction and Dutch is taught as a foreign language ($n = 18$) and others were attending bilingual English–Dutch secondary schools (*tweetalig onderwijs*) ($n = 9$).

Table 1 provides complete biographical data for all children, divided into age groups; the older children are collapsed into two groups (12 and 13 year olds and 14 to 17 year olds) to ensure that – with the exception of the 9 year olds – the number of children per age group is more or less equal. The children's scores on standardised vocabulary tests, used here as a general measure of language proficiency, are also included; the tests used were Peabody Picture Vocabulary Test 4 (Dunn & Dunn, 2007) or British Picture Vocabulary Scale (Dunn, Dunn, Whetton & Burley, 1997) for English, depending on the variety to which the child had been exposed, and PPVT-III-NL (Dunn, Dunn & Schlichting, 2005) for Dutch. The reported scores are standard scores (normed for monolinguals).

Given that all the children are simultaneous bilinguals, traditional length of exposure is the same as chronological age. The values for CUMULATIVE length of exposure are on average just over half of the traditional values, but there is considerable variation, which reflects the large range of values observed for current amount of exposure: the average group scores vary from 46% to 77%, whereas individual scores vary from 8% to 93%. The average scores for both Dutch and English vocabulary show that as a group, the children fall within the range of age-matched typically-developing monolinguals.

Data were also collected from 26 monolingual Dutch 4 to 6 year olds ($M 5.8$, $SD 0.92$). Their average score on the PPVT-III-NL vocabulary task was 109 ($SD 10.3$). Given that the age range for the bilingual children is larger than for the monolinguals, monolingual data from the most comparable study available in the literature, Blom et al. (2008a), collected using an almost identical task, will be used as a basis of comparison for the 3 and 7 year olds.

5.3 Materials

Two elicited production tasks and one grammaticality judgement task were used to collect information about children's knowledge of gender-marking on definite determiners (all tasks) and adjectives in indefinite DPs (picture description task only). In the first task, a picture description task based on Blom et al. (2008a), children

Table 1. Overview of participants.

Group	n	Length of exposure (traditional) (in years)	Length of exposure (cumulative) (in years)	% current exposure to NL	Dutch vocabulary score	English vocabulary score
3 year olds	11	3.6 <i>SD</i> 0.19	2.4 <i>SD</i> 0.42	50.3 <i>SD</i> 19.0	109 <i>SD</i> 12.3	94 <i>SD</i> 9.0
4 year olds	13	4.3 <i>SD</i> 0.31	2.4 <i>SD</i> 0.55	60.5 <i>SD</i> 14.2	103 <i>SD</i> 10.6	92 <i>SD</i> 19.1
5 year olds	14	5.4 <i>SD</i> 0.26	2.8 <i>SD</i> 0.97	55.0 <i>SD</i> 20.5	110 <i>SD</i> 15.9	99 <i>SD</i> 9.7
6 year olds	15	6.5 <i>SD</i> 0.27	3.7 <i>SD</i> 0.43	54.6 <i>SD</i> 17.4	107 <i>SD</i> 12.9	91 <i>SD</i> 8.6
7 year olds	14	7.4 <i>SD</i> 0.25	4.1 <i>SD</i> 0.88	65.1 <i>SD</i> 14.7	117 <i>SD</i> 7.4	97 <i>SD</i> 15.0
8 year olds	15	8.3 <i>SD</i> 0.28	4.2 <i>SD</i> 1.12	62.4 <i>SD</i> 23.2	109 <i>SD</i> 11.9	96 <i>SD</i> 12.2
9 year olds	8	9.5 <i>SD</i> 0.31	5.4 <i>SD</i> 0.83	76.8 <i>SD</i> 13.0	108 <i>SD</i> 16.9	92 <i>SD</i> 26.2
10 year olds	11	10.4 <i>SD</i> 0.26	5.6 <i>SD</i> 1.34	64.6 <i>SD</i> 19.1	104 <i>SD</i> 13.5	90 <i>SD</i> 11.9
11 year olds	13	11.5 <i>SD</i> 0.23	5.8 <i>SD</i> 1.73	53.9 <i>SD</i> 19.0	117 <i>SD</i> 11.5	97 <i>SD</i> 14.6
12 and 13 year olds	11	12.8 <i>SD</i> 0.39	7.3 <i>SD</i> 1.09	63.9 <i>SD</i> 15.3	113 <i>SD</i> 11.7	98 <i>SD</i> 19.5
14 to 17 year olds	11	15.1 <i>SD</i> 1.15	8.7 <i>SD</i> 2.24	45.5 <i>SD</i> 16.1	105 <i>SD</i> 13.2	93 <i>SD</i> 16.1

are presented with two pictures, e.g., a yellow and a blue robot, and asked to name them using the following prompt: “Look! Here we see two pictures. This is a . . . (child: yellow robot). And this is a . . . (child: blue robot)”. To elicit definite determiners, an additional item, e.g., a ball, subsequently appears next to each of the objects and the child is asked to complete the following prompt: “The ball is in front of . . . (child: the yellow robot). And the finger is pointing to . . . (child: the blue robot)”. Each noun is thus elicited with a definite determiner twice in this task.⁵ Fillers were items testing verb form and placement (used for another part of the same project).

A further, third definite determiner token is elicited for each noun in a story task, where children help tell a story to a puppet using pictures. Children are first asked to name the relevant nouns, and subsequently to name the same items one by one in response to a series of questions, thereby eliciting definite DPs. For example, the children

are told a story about a boy and a girl who visit the petting zoo, where they see a deer, a sheep and a rabbit. The children name each animal as it appears on the screen. They are then told that the children in the story want to feed the animal and are asked a question, such as “Which of these three animals is given a sandwich?”. A sandwich appears next to the deer and the child is expected to say “the deer”.

The grammaticality judgement task was a forced choice task using congruent and incongruent (definite) determiner–noun combinations. In order to create a felicitous context for the use of a definite DP, pictures of the relevant items were first presented and named by the experimenter (“Here we see a baby, a house, a tree, etc.”).⁶ Subsequently, each item was presented individually and two previously introduced puppets were asked to name what they saw. In doing so, one puppet used a congruent determiner–noun combination, e.g.,

⁵ Data were thus also collected for adjectival inflection in definite DPs. These are excluded from the present analysis because children were generally at ceiling in all groups and hence the results are relatively uninteresting in the present context (but see Unsworth, in press).

⁶ Recall that there is no gender-marking on indefinite determiners in Dutch and so by introducing the items in this way, the experimenter does not provide any clues as to the nouns’ gender.

*de*_{COMMON} *boom*_{COMMON} “the tree”, and the other the incongruent counterpart, e.g., **het*_{NEUTER} *boom*_{COMMON} “the tree”. Children then had to say which puppet “got it right”. Filler items ($n = 12$) were used to check whether children were able to complete the task and that they were paying attention. They either contained word order errors, i.e., *de vlinders* “the butterflies” vs. **vlinders de* “butterflies the”, determiner errors with plural nouns, i.e., *de auto's* “the cars” vs. **het auto's* “the cars”, or nonsense nouns, i.e., *de banaan* “the banana” vs. *de perg* “the perg” all of which conformed to Dutch phonotactic constraints and were produced with the common definite determiner *de*. For both filler and target items, the puppets’ responses were pre-recorded using one male and one female voice. The correct response was counterbalanced across the two puppets.

The same nine nouns per gender were used in all tasks: *baby* “baby”, *boom* “tree”, *fiets* “bicycle”, *telefoon* “telephone”, *sleutel* “key”, *klok* “clock”, *gitaar* “guitar”, *helikopter* “helicopter” and *robot* “robot” for common, and *huis* “house”, *bad* “bathtub”, *raam* “window”, *konijn* “rabbit”, *schaap* “sheep”, *vliegtuig* “aeroplane”, *hert* “deer”, *net* “net” and *eiland* “island” for neuter. These were selected from a wordlist for 4- to 6-year-old monolingual children (Damhuis, de Glopper, Boers & Kienstra, 1992); the criteria for selection were that nouns should be count, non-derived, easy to depict, and they should not be highly specific to either the home or the school environment. Because these tasks were part of a larger test battery, and because younger children have a shorter attention span, two versions of the picture description task and the grammaticality judgement task were used: one for younger children (≤ 5 years) and one for older children (> 6 years). For production (both tasks), the maximum number of items per gender was 21 for younger children and 27 for older children for definite determiners, and for adjectives, 12 for younger children and 18 for older children. For judgement, due to time/concentration constraints, each noun was tested just once. The maximum number of items per gender for younger children was thus six and for older children nine. In both production and judgement tasks, any nouns which the children did not know were excluded from analysis. Each task had two presentation orders, B being the reverse of A, and these were counter-balanced across children.

5.4 Procedure

Children were tested individually by a (near-)native speaker research assistant either at home or at school. For Dutch, children first completed the two production tasks, then the vocabulary task, and subsequently the judgement task. The English vocabulary task was administered on another day with no more than two weeks between the two languages. For the production tasks, a randomly

selected subset (approximately 10%) were cross-checked by a second tester to calculate inter-rater reliability; the Kappa statistic was very high (.96, $p < .001$) indicating almost perfect agreement (Landis & Koch, 1977).

Parents either completed the questionnaire online or (where possible) in a face-to-face or telephone interview with a research assistant. Missing or incomplete answers were followed up with a telephone call to secure the required information. The completion rate was high, at 93% (127/136).

6. Results

First, results of the two production tasks are presented in Section 6.1, followed by the judgement data in Section 6.2. The data for the two production tasks are presented together.

6.1 Elicited production data

First, accuracy scores for groups defined by age are examined in order to evaluate the data from a developmental perspective and to compare the bilingual children’s results with those of their monolingual peers. Results for determiners and adjectives are analysed separately and then compared. Individual results are considered in terms of consistency of responses and ceiling performance. Finally, regression analyses are conducted in order to determine the relative contribution of the exposure variables under consideration.

Group results and bilingual–monolingual comparisons

The accuracy scores for determiners were analysed as follows: for each child, the average percentage of correct answers was calculated by dividing the number of nouns produced with the target definite determiner (*de* for common nouns or *het* for neuter nouns) by the total number of nouns of the same gender produced with either of these determiners. For the younger children, the average number of items per child produced with a definite determiner was 18 for common and 17 for neuter nouns (max. 21), and for older children, this was 26 for both genders (max. 27). For adjectives, the accuracy scores were calculated by dividing the number of DPs containing target inflection, i.e., uninflected for neuter and inflected for common, by the total number of DPs containing adjectives either with or without inflection. The average number of items per child was the maximum for both genders for the younger ($n = 12$) and older ($n = 28$) children. There was no effect of presentation order for determiners ($t(134) = -.15, p > .05$) or adjectives ($t(134) = .91, p > .05$).

The results for common nouns are presented in Figure 1 and for neuter in Figure 2. Monolingual data for 3- and 7-year-old children from Blom et al. (2008a) are included

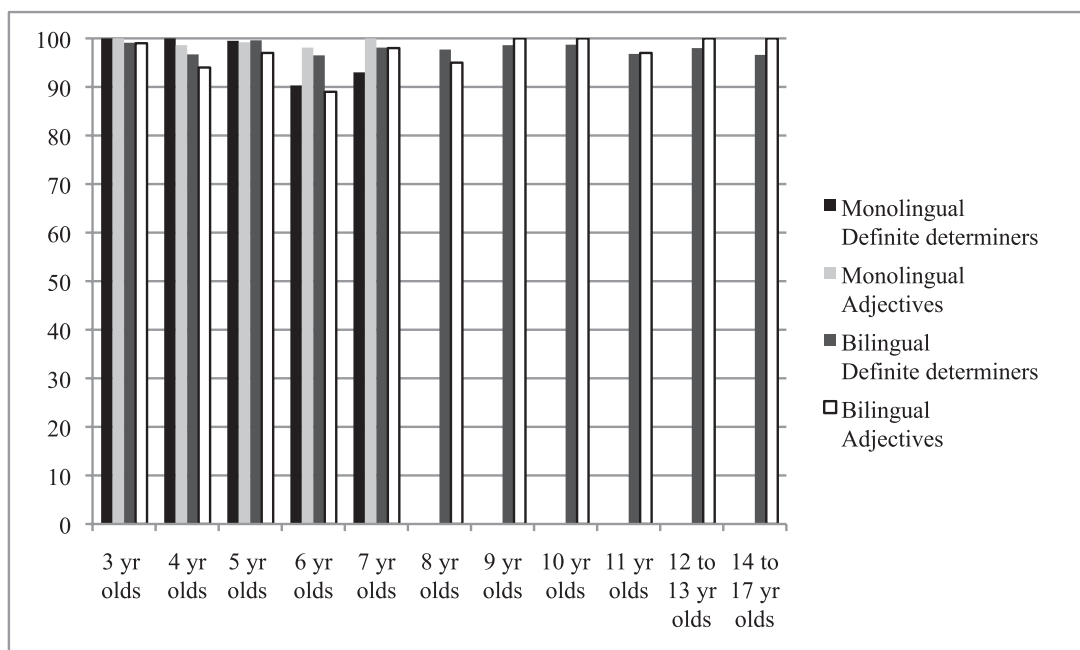


Figure 1. Average percentage of common nouns produced with target form.

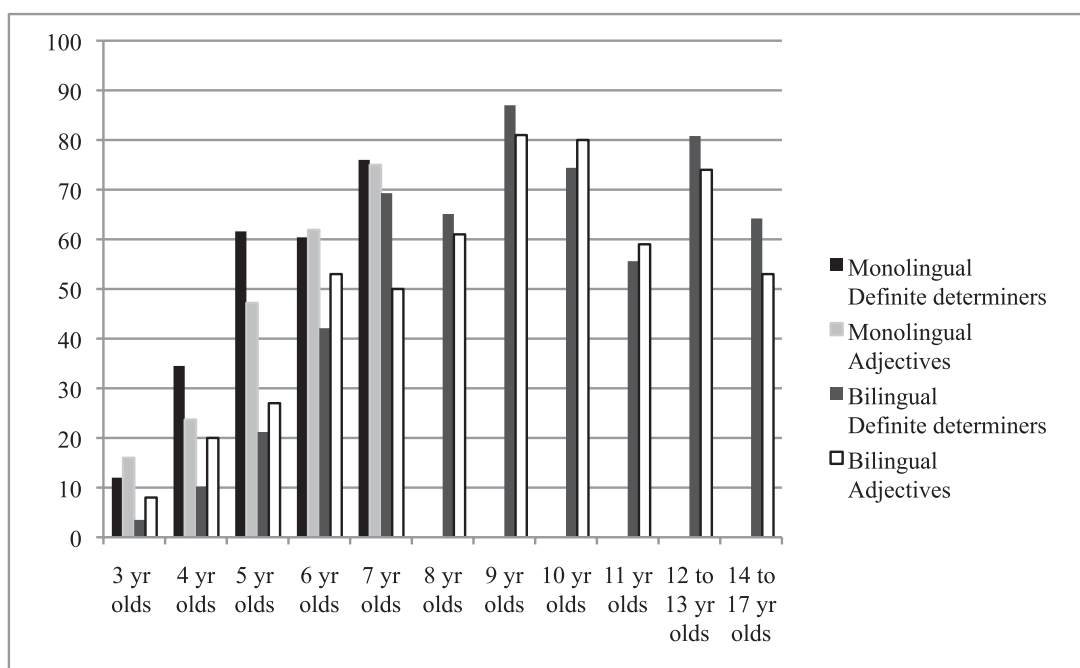


Figure 2. Average percentage of neuter nouns produced with target form (adjectives with singular indefinite nouns only, i.e., where uninflected form is expected).

for comparison.⁷ The exact data are given in Table A1 and Table A2 in the Appendix.

⁷ As already indicated, the task employed in this study is taken from Blom et al. (2008a) and is hence comparable; however, note that these authors calculated the group scores for the group as a whole rather than taking the average score across children as is the case here.

We first focus on only those bilingual children for whom we have monolingual comparison data, i.e., the 4 to 6 year olds. A mixed design ANOVA was conducted on accuracy scores on determiners with gender (common vs. neuter) as within-subjects factor and age (4 year olds vs. 5 year olds vs. 6 year olds) and group (bilingual vs. monolingual) as between-subjects factors. A significant

main effect was observed for gender ($F(1,62) = 177.0$, $p < .001$, $\eta^2_p = .74$), group ($F(1,62) = 12.7$, $p = .001$, $\eta^2_p = .17$) but not for age ($F(1,62) = 3.53$, $p > .01$).⁸ There was a significant interaction between gender and group ($F(1,62) = 10.5$, $p < .01$, $\eta^2_p = .15$), and a marginally significant interaction between gender and age ($F(2,62) = 4.55$, $p = .01$, $\eta^2_p = .13$) but no further significant interactions. In other words, averaging across groups (bilingual vs. monolingual, as well as the three different age groups), children are significantly more accurate with common than neuter nouns, both bilingual and monolingual groups' scores on neuter improve with increasing age, but the bilingual children's scores on neuter are significantly worse than the monolingual children.⁹

Turning to the results for adjectives, there was a main effect of gender ($F(1,62) = 147.5$, $p < .001$, $\eta^2_p = .70$), group ($F(1,62) = 6.54$, $p = .01$, $\eta^2_p = .10$) and age ($F(1,62) = 7.48$, $p = .001$, $\eta^2_p = .19$). Post-hoc (Bonferroni) tests revealed a significant difference between the 4 and the 6 year olds ($MD = -16\%$, $p = .001$). There was a significant interaction between gender and age ($F(2,62) = 5.50$, $p < .01$, $\eta^2_p = .15$) but no further significant interactions. Averaging across groups (bilingual vs. monolingual, as well as the three different

age groups), children are significantly better at providing the target form of the adjective for common than for neuter nouns, monolinguals are significantly better than bilinguals, and the 6 year olds are better than the 4 year olds. Furthermore, the difference between common and neuter diminishes for both groups with increasing age as scores on neuter improve.

Let us now consider the results for all the bilingual children. A mixed design ANOVA with gender as within-subjects factor and age as between-subjects factor was conducted. For determiners, there was a main effect of gender ($F(1,125) = 228.5$, $p < .001$, $\eta^2_p = .65$) and of age ($F(10,125) = 9.1$, $p < .001$, $\eta^2_p = .42$), as well as a significant interaction between the two ($F(10,125) = 8.1$, $p < .001$, $\eta^2_p = .39$). Post-hoc (Bonferroni) tests reveal significant differences (all at $p < .01$ or less) between the 3 year olds and all groups aged 7 years and older, between the 4 year olds and all groups aged 7 years and older except the 11 year olds, and between the 5 year olds, on the one hand, and the 7, 9, 10 and 12 to 13 year olds, on the other. For adjectives, there was a main effect of gender ($F(1,125) = 196.2$, $p < .001$, $\eta^2_p = .61$) and of age ($F(10,125) = 7.31$, $p < .001$, $\eta^2_p = .37$), as well as a significant interaction between the two ($F(10,125) = 4.5$, $p < .001$, $\eta^2_p = .27$). Post-hoc (Bonferroni) tests reveal significant differences (all at $p < .01$ level) between the 3 year olds and all groups 8 years and older except the 14 to 17 year olds, and between the 4 and 5 year olds, on the one hand, and the 9, 10 and 12 to 13 year olds, on the other.

Comparing bilinguals to age-matched monolinguals is the same as matching on (traditional) length of exposure. However, as the biodata in Table 1 reveal, once length of exposure is measured CUMULATIVELY, i.e., when the variation in amount of exposure inherent to a dual language setting is taken into account, the validity of this comparison is called into question (at least for the present purposes). If we take CUMULATIVE exposure, rather than chronological age, as the basis of comparison, then it turns out that the bilingual 3, 4 and 5 year olds could better be compared with a group of monolingual 2 year olds, the bilingual 6 year olds can better be compared with the monolingual 3 year olds, the bilingual 7 and 8 year olds with the monolingual 4 year olds, the bilingual 9, 10 and 11 year olds with the monolingual 5 year olds, the bilingual 12 and 13 year olds with the monolingual 7 year olds and the bilingual 14 to 17 year olds with a group of monolingual 8 year olds. Where the appropriate groups are available for statistical testing, no significant differences were found for determiners between the bilinguals and the monolinguals matched on CUMULATIVE length of exposure (LoE), but for adjectives, there was a significant difference between the bilingual 7 and 8 year olds compared (separately) with the monolingual 4 year olds ($p < .025$ for both – Bonferroni correction

⁸ Because the assumption of homogeneity was violated and because transforming the dependent variable did not substantially address this problem, a more stringent α level of .01 was adopted (as suggested by Tabachnick & Fidell, 2007).

⁹ An anonymous reviewer suggests that children's poor scores on neuter might be due to the fact that many of the selected nouns are cognates, something which – as the same reviewer acknowledges – is probably unavoidable when looking at Dutch–English bilinguals. The reviewer suggests that e.g., *huis* will also activate *house* and consequently, due to the cognate status of the common definite determiner *de* and the English definite determiner *the*, children will be more likely to produce the non-target determiner. This does not appear to be the case, however. While there is a (marginally) significant difference in accuracy scores for cognate and non-cognate neuter nouns for the 6 year olds ($t(14) = 2.15$, $p = .05$) and the 11 year olds ($t(12) = 2.62$, $p < .05$), it is the scores on the non-cognate nouns which are significantly lower (compare 38% and 46% for non-cognates and cognates, respectively, for the 6 year olds, and for the 11 year olds, 52% and 58%). Thus, while the cognate status of certain nouns may affect children's performance, this is not the case for all children, and it is not in the direction suggested. The same reviewer also suggests that as a result of English making a gender distinction on pronouns for animate but not for inanimate nouns, bilingual English–Dutch children may also make a distinction between these two semantic classes when it comes to gender-marking in Dutch. There is indeed an overall significant difference between animate and inanimate neuter nouns ($t(135) = 6.82$, $p < .001$) with higher scores for inanimate ($n = 6$) rather than animate ($n = 3$) neuter nouns (compare 54% *SD* 43% vs. 43% *SD* 42%). It may be the case that the availability of gender-marking on pronouns in English somehow hinders the acquisition of gender-marking on definite determiners in Dutch, but it may also be the case that children are simply worse on the three neuter nouns we used which happen to be animate. To fully examine this question, more data is necessary.

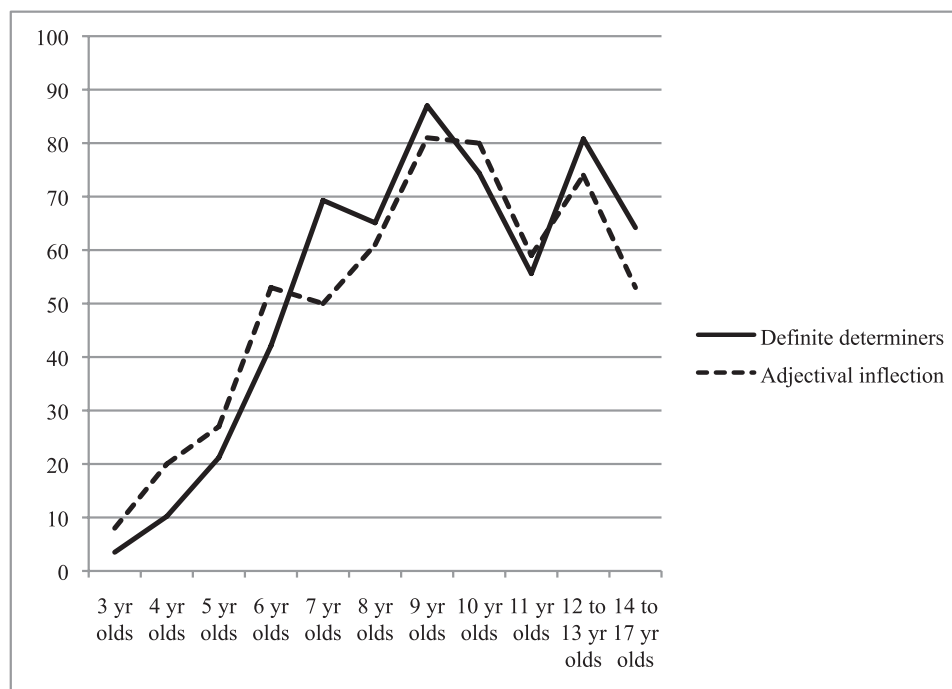


Figure 3. Production data: Average percentage correct for definite determiners and adjectives (in singular indefinite DPs) for neuter nouns (bilingual children only).

applied), and between the bilingual 9 and 10 year olds compared (separately) with the monolingual 5 year olds ($p < .01$ for both; compare bilingual 11 year olds vs. monolingual 5 year olds, $p > .017$). In all cases, the bilingual groups outperformed the CUMULATIVE-LoE-matched monolinguals.

Definite determiners vs. adjectives

Each of the bilingual and monolingual groups is at ceiling for both determiners and adjectives with common nouns. For neuter nouns, this is clearly not the case. In order to compare the bilingual children's results for determiners and adjectives directly, accuracy scores for each are plotted against each other in Figure 3.

A mixed design ANOVA with domain (determiners vs. adjectives) as within-subjects factor and age as between-subjects factor revealed no main effect of domain ($F(1,125) = .13$, $p > .01$; see fn. 8 above). In other words, as Figure 3 suggests, there seems to be a tight relation between the bilingual children's accuracy rates for determiners and adjectives on neuter nouns ($r(136) = .81$, $p < .001$), as is the case for monolingual children ($r(26) = .84$, $p < .001$).

Consistency

In order to better understand the variation on neuter nouns, further analyses are conducted of the children's individual data. First, we consider the consistency with which children use the target definite determiner with one

and the same noun, and then we use this information to reanalyse children's performance on adjectives.

Recall that a maximum of three tokens were elicited per noun. Consistent gender-marking was operationalised as either 2/2, 2/3 or 3/3 correct.¹⁰ These data are presented in Table A3 in the Appendix. The results of this analysis are in line with the accuracy rates presented above, but note that the number of children producing any consistently-marked neuter nouns is low for the youngest bilingual groups, i.e., no 3 year olds, one 4 year old and five 5 year olds.

The consistency data for determiners are now used to reanalyse children's responses on adjectives. It is after all possible that children may know the rules for adjectives but they have misattributed gender to a particular noun with the result that in terms of the target system, their response is incorrect, whereas in terms of their own system, it is perfectly accurate. Children's responses presented above in Figures 1 and 2 were thus reanalysed such that only responses for those nouns marked consistently with the target determiner were included. The results for neuter nouns are presented in Figure 4. The exact data (for both genders) are given in

¹⁰ Adopting a stricter definition of consistent (as in Blom et al. 2008a, where 2/3 is counted as inconsistent) does not alter the results in any significant way. Thus, the less stringent criterion is adopted here as it allows us to include more data and therefore increase the power of the analysis.

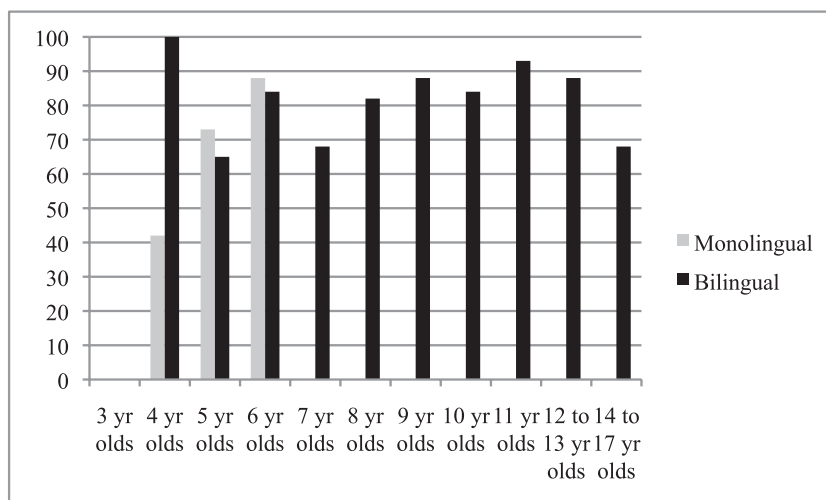


Figure 4. Reanalysis of adjective data taking into account consistency: Average percentage of consistently neuter-marked nouns produced with target adjective (in singular indefinite DPs only).

Table A4 in the Appendix. Note that the value for the 4 year olds are from one child only.

This reanalysis leads to a clear improvement in children’s scores (recall Figure 2).

Ceiling performance

As a group, only the 9-year-old bilinguals are approaching target on both determiners and adjectives. The high SDs for all groups suggest that there may however be individual children in several groups who are at target. Following Montrul, Foote and Perpiñán (2008), 90% correct was adopted as a criterion for targetlike performance. Table 2 presents an overview of the number of children in each group who meet this criterion for determiners and for adjectives.

For determiners with neuter nouns, there are no target children in the youngest three (3- to 5-year-old) groups, whereas there are some target children in all of the other groups. The relative distribution of the target children for both determiners and adjectives (for all nouns and for consistent nouns only) again reflects the accuracy results, with proportionally more children in the 9-year-old and 12-and-13-year-old groups reaching target.

Regression analysis

In order to determine the relative contribution of exposure and proficiency, a multiple linear backward-elimination regression analysis was conducted. Chronological age was excluded because it correlated strongly with CUMULATIVE length of exposure ($r(127) = .86, p < .001$). The results are given in Table 3.

For determiners, all three variables contributed significantly to the model, with CUMULATIVE length of exposure accounting for the most variance. The beta

values can be interpreted as follows: holding the effects of the other two variables constant, for every year of CUMULATIVE exposure, there is a 10% increase in children’s accuracy on definite determiners with neuter nouns, for every additional 10% current exposure to Dutch, there is a 7% increase in accuracy scores, and for every one point increase on the PPVT-III-NL, there is a 4% increase in accuracy scores.

Given that children’s performance on adjectives was shown to be related to their performance on determiners, accuracy scores on determiners were also included in the regression analysis for adjectives. This allows us to evaluate whether the exposure/proficiency variables contribute to children’s accuracy on adjectives over and above the contribution each of these variables make to children’s accuracy on determiners.¹¹ The results were as follows: in the final model ($R^2 = .66, F(2,123) = 120.0, p < .001$), the only significant predictor variable was children’s accuracy scores on determiners ($\beta = .85, p < .001$).

Exposure patterns in the early years

Whereas the younger children’s results show a clear developmental trend, i.e., scores improve with age, this trend discontinues at around age eight. This may be due to the specific characteristics of our sample or it may be a property of the language acquisition process. In this section, we examine this issue by determining whether the (non-target) older children’s past exposure patterns are significantly different from those of the younger children.

¹¹ In order to meet the assumption of normally distributed errors (Field, 2009, p. 221), the data were first transformed using a logarithmic function.

Table 2. Percentage (and number) of children at target (i.e., $\geq 90\%$) for production of gender-marking on definite determiners and adjectives.

Group	n	Definite determiners		Adjectives		Adjectives (consistent Ns only)	
		Common	Neuter	Common	Neuter	Common	Neuter
Monolinguals							
4 year olds	6	100% (6)	0% (0)	100% (6)	0% (0)	100% (6)	17% (1)
5 year olds	11	100% (11)	18% (2)	100% (11)	9% (1)	100% (11)	28% (3)
6 year olds	9	67% (6)	22% (2)	89% (8)	11% (1)	89% (8)	44% (4)
Bilinguals							
3 year olds	11	100% (11)	0% (0)	100% (11)	0% (0)	91% (10)	0% (0)
4 year olds	13	92% (12)	0% (0)	92% (12)	8% (1)	92% (12)	8% (1)
5 year olds	14	100% (14)	0% (0)	93% (13)	0% (0)	93% (13)	14% (2)
6 year olds	15	87% (13)	13% (2)	67% (10)	13% (2)	85% (11)	40% (6)
7 year olds	14	100% (14)	29% (4)	86% (12)	14% (2)	93% (13)	36% (5)
8 year olds	15	93% (14)	47% (7)	93% (14)	27% (4)	93% (14)	40% (6)
9 year olds	8	100% (8)	75% (6)	100% (8)	50% (4)	100% (8)	63% (5)
10 year olds	11	100% (11)	55% (6)	100% (11)	28% (3)	91% (10)	55% (6)
11 year olds	13	92% (12)	38% (5)	77% (10)	31% (4)	85% (11)	38% (5)
12 and 13 year olds	11	91% (10)	73% (8)	100% (11)	55% (6)	100% (11)	64% (7)
14 to 17 year olds	11	91% (10)	45% (5)	100% (11)	36% (4)	100% (11)	55% (6)

Table 3. Results of regression analysis for production of definite determiners for neuter nouns.

Factor	Unstandardised coefficients		Standardised coefficients
	B	St. error	β
(constant)	-.85	.23	
CUMULATIVE length of exposure	.10	.01	.53***
Current amount of exposure	.70	.15	.31***
Vocabulary score	.004	.002	.14*

adj. $R^2 = .50$ ($F(3,122) = 42.0, p < .001$); * $p < .05$, *** $p < .001$

Recall that the parental questionnaire includes information about children's language exposure at daycare, school and home over time, for each one year period in the child's life thus far. Instead of summing this information as in preceding sections, here we examine the data for the early years separately to determine whether there are any differences in early exposure patterns which might explain the observed differences in (the inferred) developmental trajectory between younger (3 to 7 year olds) and older (8 to 17 year olds).

The results are given as the average proportion of a given year with exposure to Dutch, for each year from birth to age 7 years in Figure 5.

There is a significant difference between the older and the younger children for the periods from birth to age 1 ($t(107) = 2.15, p < .05, d = .42$), from age 1 to 2 ($t(107) = 2.01, p < .05, d = .39$), from age 2 to 3 ($t(107) = 2.90, p < .01, d = .56$), and from age 3 to 4 ($t(96) = 2.81, p < .01, d = .57$), but not for the periods from age 4 to 5 ($t(84) = -1.77, p > .05$), from age 5 to 6 ($t(71) = .34, p > .05$) or from age 6 to 7 ($t(56) = .35, p > .05$). From birth to age 4, the younger children have thus – as a group – had significantly more exposure to Dutch than the older children.

In order to determine whether the older children's exposure patterns in the early years can account for their accuracy on neuter nouns, the regression analyses (for both determiners and adjectives) conducted above were repeated and the total amount of exposure from birth to age 4, i.e., the periods for which a significant difference was observed between the younger and older children, was included as a predictor variable. This did not change the results. To check whether the relevant period may be longer, i.e., from birth to 6 years, the analysis was repeated for the older children with exposure in the early years from birth to age 6 included as a predictor: once again, this did not change the results.

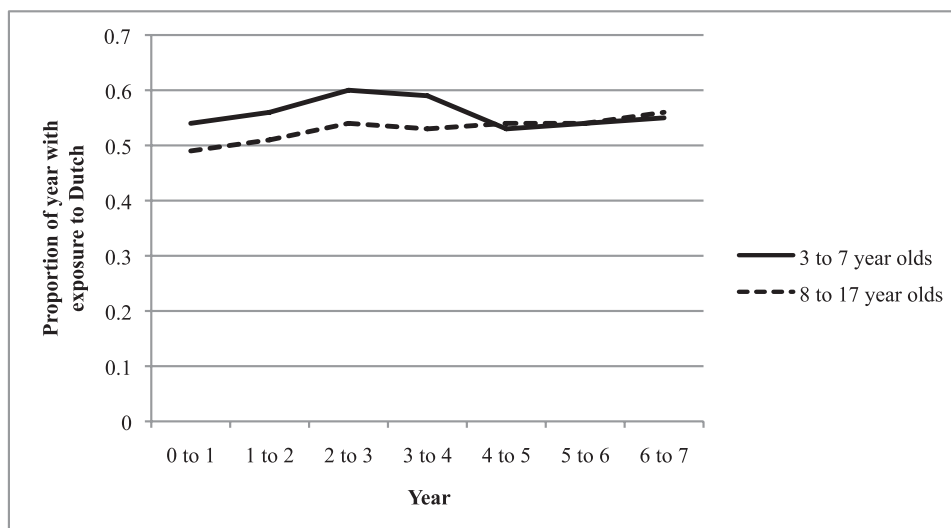


Figure 5. Language exposure patterns in early years.

Current exposure patterns

The regression analysis indicated that the current amount of Dutch (at home and school) to which children are exposed is a significant predictor of their accuracy scores on definite determiners. A number of the children in our sample attend (predominantly) English-language or bilingual schools. It is possible that this may contribute to some of the other children's poor performance on definite determiners. To explore this possibility, the older children were divided into two groups: those who attend Dutch-language schools ($n=48$) and those who attend bilingual or English-language schools ($n=20$). An independent t -test ($t(28)=-4.61, p < .001, d=1.30$) revealed that scores for the latter group (38% SD 39%) were significantly lower than for the former (83% SD 28%). Note, however, that these two groups do not differ on CUMULATIVE length of exposure ($t(24)=-.13, p > .05$).

Summary of elicited production results

Consonant with previous findings, both monolingual and bilingual children were more accurate on common than neuter nouns. Where the relevant data were available, bilingual children were generally less accurate than their monolingual peers, although when matched on CUMULATIVE length of exposure, this discrepancy disappeared. For both monolingual and bilingual children, there is a close relation between accuracy scores on definite determiners and adjectives. Accuracy on adjectives with neuter nouns improves when gender attribution is taken into account. Individual results are in line with the group data. CUMULATIVE length of exposure, current amount of exposure and vocabulary score are all significant predictors of the bilingual children's scores

on definite determiners with neuter nouns, with the first of these accounting for the most variance. The only significant predictor for accuracy scores on adjectives is children's scores on determiners with the same nouns. Children's exposure patterns in the early years (birth to age 4 or birth to age 6) were not a significant predictor variable for older children's accuracy scores, and older children attending English-language or bilingual schools scored significantly lower on definite determiners.

6.2 Grammaticality judgement data

Three children were unable to complete the task. Fillers were used to exclude children with a puppet bias ($n=5$) or who appeared to be randomly selecting a puppet ($n=10$). Given that most of these children were 3 and 4 year olds, thereby significantly reducing the numbers for these two groups, the analysis concerns children aged 5 and older only. The analysis of the judgement data follows the same steps as for production.

Group results and bilingual–monolingual comparisons

The accuracy scores were analysed as follows: for each child, the average percentage of correct answers was calculated by dividing the number of nouns for which the child selected the congruent determiner–noun combination by the total number of items of the same gender to which the child responded. There was no effect of presentation order ($t(107)=.33, p > .05$). The results for common and neuter nouns are presented in Figure 6. The accuracy scores for the monolingual children were for the 5 year olds 88% (SD 20%) for common and 63% (SD 34%) for neuter, and for the 6 year olds 93% (SD 6%)

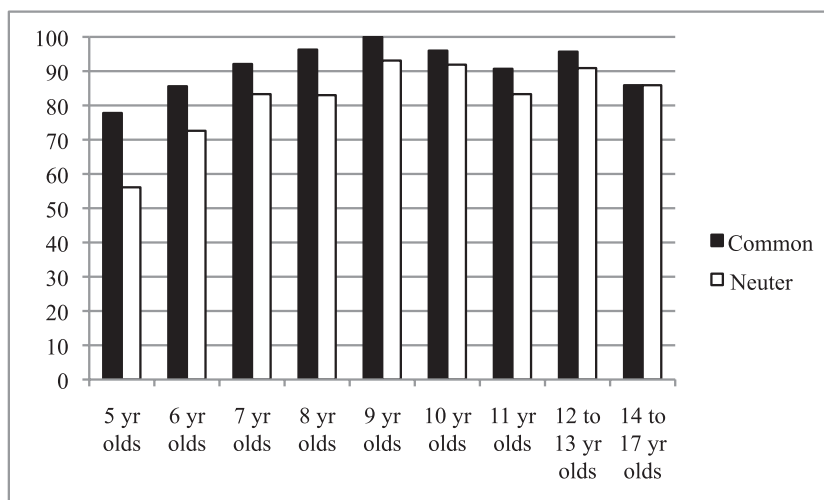


Figure 6. Average percentage of nouns selected in grammaticality judgement task with target determiner.

for common and 78% (*SD* 18%) for neuter. The exact data are given in Table A5 in the Appendix.

In this task, children are forced to choose between two items; to check whether each group’s performance is significantly different from chance, a one sample *t*-test was conducted for the groups separately with the test value set at 50% and the alpha corrected accordingly. All monolingual and bilingual groups were significantly different from chance ($p < .01$ or lower) on both common and neuter nouns with the exception of the 5 year olds (bilingual and monolingual) who were at chance level for neuter nouns.

As for the elicited production data, a mixed design ANOVA was first conducted with group (bilingual vs. monolingual) and age (5 year olds vs. 6 year olds) as between-subjects factors and gender (common vs. neuter) as within-subjects factor. There was a main effect of gender ($F(1,42) = 10.5, p < .01, \eta^2_p = .20$), and of age ($F(1,42) = 4.68, p < .05, \eta^2_p = .10$), but not of group ($F(1,42) = 2.28, p > .05$). There were no significant interactions.

Turning now to the whole bilingual dataset, a mixed design ANOVA with gender as within-subjects factor and age as between-subjects factor was conducted. The results were as follows: there was a main effect of gender ($F(1,100) = 12.8, p = .001, \eta^2_p = .11$) and of age ($F(8,100) = 3.8, p = .001, \eta^2_p = .23$) but no interaction between the two ($F(8,100) = .73, p > .01$; see fn. 8 above). Post-hoc (Bonferroni) tests indicate a significant difference (at $p < .01$ or lower) between the 5 year olds, on the one hand, and the 8, 9, 10, 12 and 13 year olds, on the other. There were no further between-group differences.

Once again, if we compare the bilingual children with available monolingual data when matched on CUMULATIVE length of exposure, where relevant data is available, i.e., for the bilingual 9, 10 and 11 year olds with

Table 4. Percentage (and number) of children at target (i.e., $\geq 90%$) for judgement of gender-marking on definite determiners.

Group	n	Common	Neuter	Both
Monolinguals				
5 year olds	10	90% (9)	40% (4)	30% (3)
6 year olds	9	100% (9)	44% (4)	44% (4)
Bilinguals				
5 year olds	14	57% (8)	29% (4)	14% (2)
6 year olds	15	73% (11)	33% (5)	27% (4)
7 year olds	14	93% (13)	71% (10)	64% (9)
8 year olds	15	93% (14)	67% (10)	67% (10)
9 year olds	8	100% (8)	88% (7)	88% (7)
10 year olds	11	91% (10)	91% (10)	82% (9)
11 year olds	13	69% (9)	62% (8)	62% (8)
12 and 13 year olds	11	91% (10)	82% (9)	82% (9)
14 to 17 year olds	11	64% (7)	64% (7)	55% (6)

the monolingual 5 year olds, no significant differences are observed between groups ($p > .017$ for all comparisons (with Bonferroni correction)).

Ceiling performance

Given that each noun is judged only once – hence no consistency analysis for these data – a 90% criterion was considered too strict because in order to reach target, a child would have to judge all nouns correctly (i.e., 6/6 for the younger children and 9/9 for the older children). Thus, allowing room for noise as we did for the production data, the criterion of 5/6 or 8/9 correct was adopted. The results are given in Table 4.

Table 5. Results of regression analysis for grammaticality judgement task for common nouns.

Factor	Unstandardised coefficients		Standardised coefficients
	B	St. error	β
(constant)	.16	.02	
CUMULATIVE length of exposure	.006	.003	.20*
Current amount of exposure	.16	.03	.51***

Note: adj. $R^2 = .34$ ($F(2,97) = 26.1, p < .001$); * $p < .05$, *** $p < .001$

Table 6. Results of regression analysis for grammaticality judgement task for neuter nouns.

Factor	Unstandardised coefficients		Standardised coefficients
	B	St. error	β
(constant)	.47	.07	
CUMULATIVE length of exposure	.04	.01	.36***
Current amount of exposure	.27	.11	.24*

adj. $R^2 = .22$ ($F(2,97) = 14.8, p < .001$); * $p < .05$, *** $p < .001$

As with the production data, there are more target children for common nouns than for neuter, but there are also a number of non-target children for common (recall production – see also Section 6.3). For neuter nouns, approximately a third of the 5 and 6 year olds reach target and in all other groups, at least around two-thirds of the children do so. With a handful of exceptions, children who reach target on neuter nouns are also target on common nouns, but this does not always hold the other way.

Regression analyses

As for the production task, all independent variables with a significant bivariate correlation with the dependent variable were entered into a backward-elimination regression analysis. For common nouns, CUMULATIVE length of exposure, current amount of exposure and vocabulary score, and for neuter nouns, only the first two predictor variables were included (see fn. 11 above).

Both CUMULATIVE length of exposure and current amount of exposure are significant predictor variables for both common and neuter nouns. The standardised

coefficients indicate that current amount of exposure accounts for more of the variance with common nouns, whereas this pattern is reversed for neuter nouns although the difference between the two predictor variables is not as large.

Current exposure patterns

As for the production data, current exposure to Dutch was also found to be a significant predictor for the judgement data. Once again we find children at Dutch-language schools score higher (neuter: 77% *SD* 25%; common: 96% *SD* 12%) than those at English-language or bilingual schools (neuter: 92% *SD* 14%; common: 87% *SD* 15%; neuter: $t(25) = -2.58, p < .05, d = .75$; common $t(29) = -2.35, p < .05, d = .66$).

Summary of grammaticality judgement results

As for production, children are significantly more accurate for common than neuter nouns. There was however quite some variation for common as well as for neuter gender. Where monolingual comparison data were available, no significant differences were found between bilinguals and monolinguals; the 5 year olds were significantly less accurate than the older bilingual children. Individual response patterns are in line with the group results. Both exposure variables were found to be significant predictor variables for common and neuter nouns, albeit to differing degrees.

6.3 Elicited production and grammaticality judgement data compared

Recall that the production data were elicited using two similar tasks. In the picture description task, each noun was elicited with a definite determiner alongside an adjective (see fn. 5 above), and in the story task, the definite determiner was elicited by itself. In order to compare children’s performance on the production and judgement tasks more precisely, the average percentage correct for the judgement data is now compared with the average percentage correct for this second task. There is thus only one token per noun and only those nouns with data in both tasks are included in the analysis.

Within-group analysis

The results for all monolingual and bilingual children who completed both tasks are presented in Figure 7.¹²

A mixed design ANOVA was conducted for the bilingual children with gender and modality as

¹² The scores in Figure 7 may differ slightly from those presented in preceding sections because they contain only those children who completed both tasks and because data from the picture description task are excluded (see Unsworth et al., in press, for discussion of possible between-task differences).

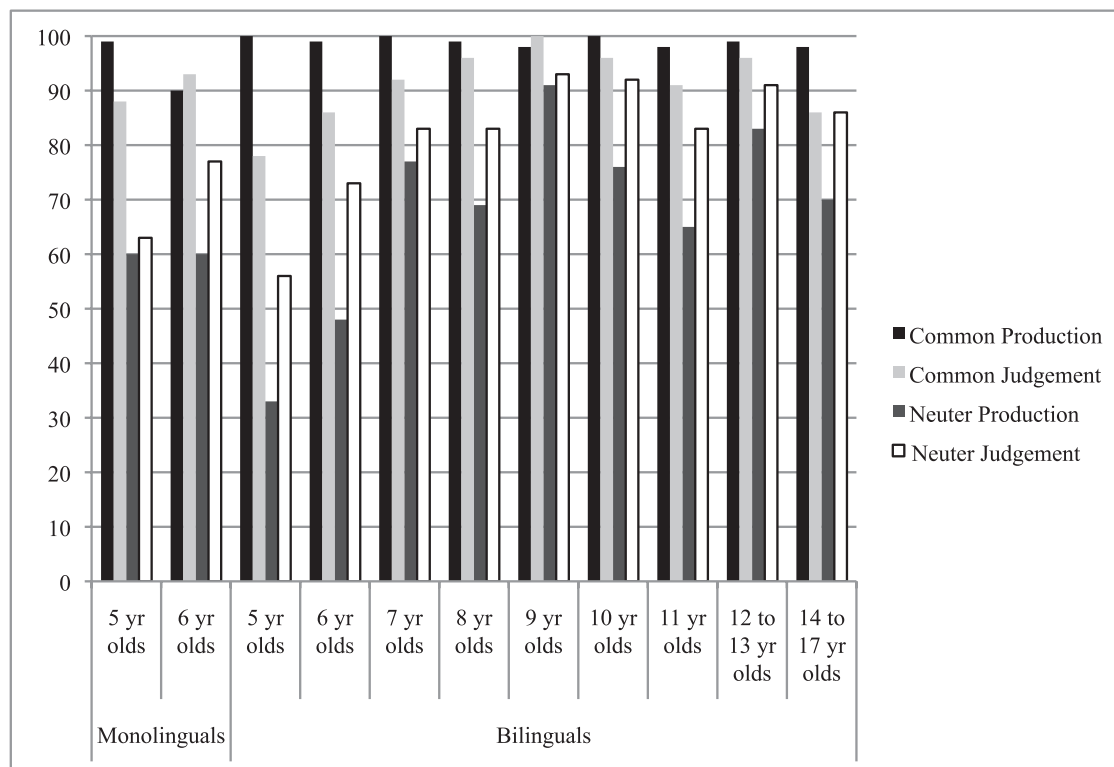


Figure 7. Average percentage correct on production and judgement tasks: same nouns, determiner–noun only.

within-subjects factor and age as between-subjects factor. There was a main effect of gender ($F(1,100) = 71.3$, $p < .001$, $\eta^2_p = .42$), modality ($F(1,100) = 9.56$, $p < .01$, $\eta^2_p = .09$) and of age ($F(8,100) = 3.59$, $p = .001$, $\eta^2_p = .23$), as well as significant interactions between gender and age ($F(8,100) = 2.68$, $p = .01$, $\eta^2_p = .18$), and gender and modality ($F(1,100) = 35.4$, $p < .001$, $\eta^2_p = .26$). Children are thus significantly more accurate on judgement than on production and this holds for neuter more so than for common nouns.

Response patterns across tasks per noun

In order to further compare bilingual children's behaviour on the production and judgement tasks, data were examined from individual nouns, and for each child, we calculated the proportion of nouns for which responses were (i) target on both tasks, (ii) target on production but not on judgement, and (iii) target on judgement but not on production. The fourth logically possible pattern, i.e., target on neither, did not occur. The results are presented in Figure 8 for common nouns and Figure 9 for neuter nouns.

For most common and neuter nouns, children are at target on both tasks. For the remainder, response patterns for the two genders differ: for common gender, virtually all nouns are target on production but non-target on judgement, whereas for neuter gender, the existence of this

pattern is negligible and the reverse pattern predominates, i.e., target on judgement but non-target on production. This overall distribution holds across all groups.

Summary of production and judgement compared

When children's accuracy scores on judgement and production were compared, an interesting asymmetry emerged between common and neuter nouns: whilst common nouns were more likely to be target on production but not on judgement, the reverse pattern held for neuter nouns.

7. Discussion

In this paper we examined data on the acquisition of gender-marking on definite determiners and adjectives in indefinite DPs by simultaneous English–Dutch bilingual children to investigate the effect of current and previous amount of exposure, and to determine whether bilingual children are able to acquire the relevant abstract grammatical features and rules and apply them consistently.

7.1 Definite determiners vs. adjectives

The first research question (see Section 4 above) asked what the effect was of differential amounts of exposure

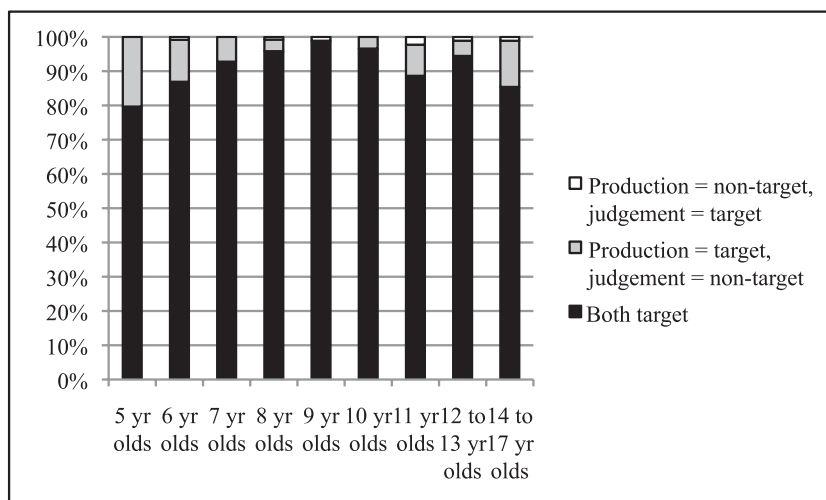


Figure 8. Average proportion of common nouns with given response pattern in production and judgement tasks, determiner-noun only.

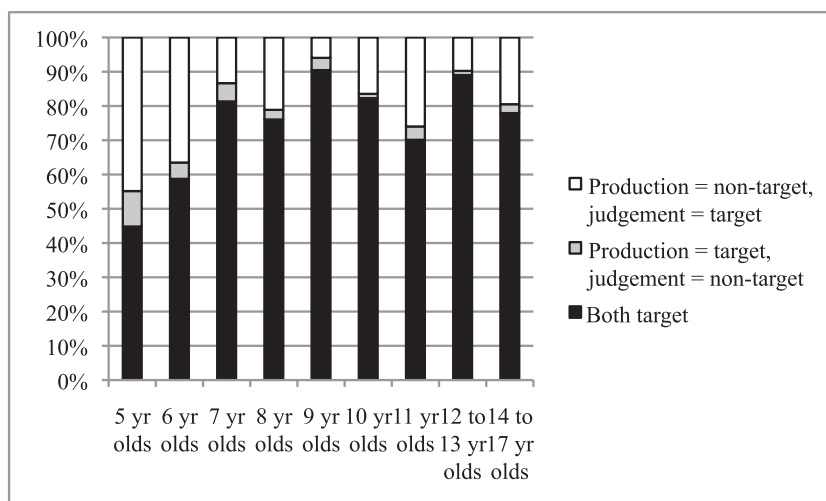


Figure 9. Average proportion of neuter nouns with given response pattern in production and judgement tasks, determiner-noun only.

– now and in the past – on the acquisition of Dutch gender. A number of predictions were made. First, it was predicted that there should be a significant effect of amount of exposure on children’s gender-marking on determiners, and that once matched on CUMULATIVE length of exposure, there should be no differences between bilinguals and monolinguals. The results confirm both predictions, for both production and judgement. CUMULATIVE length of exposure and current amount of exposure accounted for half of the variance in scores on determiners with neuter nouns in the production task (in combination with vocabulary scores), and approximately a third of the variance in scores on the judgement task. Furthermore, when bilingual children are compared

with the best-matched monolingual group in terms of CUMULATIVE length of exposure, the differences observed in the age-based bilingual–monolingual comparisons disappear; the bilingual children’s scores are as high as (or higher than) the monolinguals’.

A further prediction was that on a rule-based approach, where gender-marking on adjectives results from the application of lexical insertion rules which make use of abstract grammatical features (Blom et al., 2008a), exposure effects should be restricted to definite determiners. The results are indeed consistent with this approach: although both exposure variables correlated significantly with children’s accuracy scores on adjectives, this relationship was mediated by their scores on

determiners. The observation that amount of exposure affects gender attribution (definite determiners) more than gender agreement (adjectival inflection) is in line with recent work on simultaneous German–French and Italian–German bilinguals by Bianchi (in press) and Stöhr, Akpinar, Bianchi and Kupisch (2012).

The final prediction with respect to the first research question concerned the piecemeal approach to the acquisition of opaque gender systems, put forward for the acquisition of (Welsh), by Gathercole and Thomas (2005, 2009) and Thomas and Gathercole (2007). It was predicted that, assuming that like Welsh, Dutch has an opaque gender system, exposure effects should be found across the board, i.e. for gender-marking on both definite determiners and adjectives, and that for children with comparatively little exposure, acquisition may be “timed off the map”. The observation that amount of exposure (CUMULATIVELY and at the current time) account for approximately half of the variance in children’s scores on determiners on the production task is consistent with this prediction; however, as noted above, although significantly correlated with exposure variables, children’s scores on adjectives are best predicted by their scores on determiners only.

Further evidence for the claim that children’s responses reflect rule-based knowledge comes from the observation of a tight relation between scores on determiners and adjectives within each age group (recall Figure 3 above), as observed for monolingual L1 children by Polišenská (2010). If bilingual children’s acquisition (initially) proceeds in a piecemeal fashion, it is not clear why such a link should pertain across all groups, and especially in the younger groups, as this constitutes evidence for a rule-based system which employs notions such as abstract gender features, as in [+neuter]. Furthermore, the fact that children make errors in one direction only and do not simply reproduce what they hear in the input suggests that they have abstract and input-independent representations.¹³

The motivation behind matching bilinguals and monolinguals based on CUMULATIVE length of exposure is to illustrate an alternative, potentially more informative approach to straightforward age-based bilingual–monolingual comparisons (see Paradis, 2010b, for relevant discussion). It is freely acknowledged, however, that such comparisons are considerably more complex than the rather simplistic fashion in which they are presented here. Furthermore, they are only as good as the parental questionnaire data upon which they are based. Further research is necessary to test the applicability of the notion of CUMULATIVE length of exposure to other domains and learners.

¹³ I thank an anonymous reviewer for pointing this out.

7.2 *Nativelike ultimate attainment?*

Examining data from older simultaneous bilingual children allows us to say something about ultimate attainment in addition to development. An analysis of children’s individual response patterns showed that almost one third were at ceiling on determiners for both common and neuter on the production task, and once consistency was taken into account, the approximate number of children reaching ceiling for adjectives was similar. On the judgement task, almost half of the children were at ceiling on both genders.

Given that the sample of children included younger children whose gender systems were still developing as well as those who may be considered to have reached ultimate attainment, the existence of non-target children is unsurprising. Failure to reach target was not restricted to these younger children, however. Two possible explanations were explored for children’s errors: timing and amount of exposure, i.e., amount of exposure in the early years, and modality.

7.3 *Exposure patterns in the early years*

It was predicted that children’s poor performance on adjectives may be due to a failure to reach the relevant threshold to acquire the rule in question in the early years as a consequence of reduced exposure. It turned out that the older (8- to 17-year-old) children in our sample were estimated as having significantly less exposure to Dutch in the first four years of life than the younger (3- to 7-year-old) children, suggesting that this might contribute to the non-target behaviour observed in this group; however, when included in the regression analysis, amount of exposure in the early years, either from birth to age four or to age six, did not turn out to be a significant predictor of children’s accuracy scores on adjectives.¹⁴ This suggests that if there is a certain threshold to be met in order to acquire the relevant lexical insertion rule for adjectival inflection in Dutch, as speculated by Blom et al. (2008a), these children have reached it. More generally, it may indicate that it is amount of exposure in general and not amount of exposure in the early years which is the relevant variable here.

¹⁴ The observation that the older children had significantly less exposure in the early years could of course reflect a methodological artefact, i.e., we cannot rule out that the parents of older children somehow completed the questionnaire differently from the parents of younger children due to a greater amount of time having elapsed between the period in question and the moment at which the questionnaire was completed. It is not clear, however, why this should lead parents to systematically under-estimate (rather than over-estimate) their children’s exposure to Dutch. Furthermore, a study by Gilger (1992) suggests that retrospective parental report is not adversely affected by the children’s age.

Further evidence for this interpretation of the present findings comes from the existence of successive bilingual children with target accuracy rates on adjectival inflection (Unsworth, in press); these children will by definition not have had any target language exposure in (at least) the first four years of life. A difference in overall amount of exposure is also the likely explanation for the generally more accurate scores for the children in the present study when compared with bilingual/L2 children in previous studies (e.g., Cornips et al., 2006), although in many of these, it is possible that exposure to a variety of Dutch which is characterised by gender errors (also) contributes to children's lower accuracy scores (Blom & Vasic, 2011; Cornips & Hulk, 2008).

Additional post-hoc analyses of the older children's responses based on the language of schooling revealed that for both production and judgement, children who attended an English-language or bilingual school at the time of testing had significantly lower scores than those who attended a Dutch-language school. These findings underscore the importance of continual use and exposure for a target language property such as gender-marking on definite determiners in Dutch. This is in line with previous studies which have emphasised the role of input and children's own language use in the acquisition of gender by simultaneous bilinguals and early successive bilinguals/heritage speakers (Gathercole & Thomas, 2009; Montrul et al., 2008; Stöhr et al., 2012). Note, however, that unlike some of these studies, which claim that simultaneous bilingual children should reach the same level as monolingual children in the majority language, i.e., the language of the community in which they are growing up (Bianchi, in press; Stöhr et al., 2012), the present findings suggest that for opaque gender systems this may not be the case for some children, and especially those who are not (solely) educated in the majority language.

Recent results on English–Greek bilinguals furthermore show that when the target language is relatively systematic and transparent in its gender-marking, 2L1 children are at ceiling in a similar timeframe to L1 children (Unsworth et al., in press). The acquisition of gender in Dutch may be seen as comparable to the acquisition of gender for nouns without morphophonological cues in languages such as Spanish; it is in fact such nouns which were used in Gathercole's (2002a) study, in which exposure effects for bilinguals were observed.

An alternative view of the learning task presented in this paper, as suggested by an anonymous reviewer, is as the acquisition of a default (*de*) rule with sets of exceptions. On this view, the acquisition of other linguistic phenomena presenting a similar learning profile should be subject to the same exposure effects, e.g., the English comparative.

7.4 Production vs. judgement

Our final prediction concerning bilingual children's ability to acquire and use the abstract features and rules of the Dutch gender system was that, in line with the MSIH (Missing Surface Inflection Hypothesis; Haznedar & Schwartz, 1997; Prévost & White, 2000), (some) bilingual children's failure to consistently produce target forms may reflect a production-specific performance problem rather than a failure to acquire those grammatical features and rules and/or to specify certain nouns with the target gender feature, and consequently, they should be more accurate on a non-production task.

The results of the grammaticality judgement task were consistent with the MSIH, i.e., children were significantly better at selecting the target determiner–noun combination than they were at producing this WITH THE SAME NOUN, at least as far as neuter gender was concerned. For common nouns, if children responded differently on the two tasks, they were target on production and non-target on judgement. This finding is also in line with the MSIH in the sense that children's use of the common definite determiner *de* in production may reflect the use of a default or least specified form (Blom & Vasic, 2011; Unsworth & Hulk, 2009; Weerman et al., 2011; see also fn. 4 above). The only responses inconsistent with the MSIH are those neuter nouns where children are target on production but not on judgement; these however constitute at most on average 10% of neuter nouns, to the extent that they occur at all.

An alternative explanation for bilingual children's significantly better performance on the judgement task could be that they are using explicit, learned knowledge about gender. In other words, responses on this task may (in part) reflect the result of explicit learning rather than the acquisition of abstract linguistic knowledge, or it may be a more general task effect. While this may of course be possible, it is not clear how this should lead to the differences we see between the judgement and production tasks for common vs. neuter nouns. The application of learned determiner–noun pairings may explain the better performance on judgement for neuter nouns, but it is difficult to see how this would account for the existence of the reverse pattern for common nouns. Nevertheless, in order to fully understand the nature of children's (developing) knowledge of Dutch gender, it would be insightful to use a test battery which includes online as well as offline measures of comprehension alongside production, as has been conducted for adult L2 Spanish by Grüter, Lew-Williams and Fernald (2012).

Even though children were generally much better on judgement than on production, there still remained a number of children in each age group who failed to reach ceiling on both common and neuter nouns on the

judgement task. This may be because these children have not had sufficient exposure to the nouns in question to specify their gender, although it must be admitted that these nouns are unlikely to be infrequent in the input to (young) children. In order to fully investigate the nature of these errors, and how children's knowledge of grammatical gender changes over time, a longitudinal study using a variety of tasks which target a larger number of nouns of varying frequencies is needed. Longitudinal data would furthermore be very informative with respect to the role of continuity of exposure. The amount of input to which a child is exposed interacts with and to a certain extent is determined by a number of factors, including for example the social context in which the languages are acquired (majority/minority, prestigious or not), schooling, and the age at which literacy is acquired. Some of these factors will remain constant throughout a child's life whereas others may vary.

Finally, children were not tested on adjectives in the judgement task, and hence we cannot say for sure whether the cause of their inaccuracies with adjectives may also be a production-specific problem resulting in use of a default, or whether they have failed to acquire the topmost rule given in (3) above. The observation that once corrected for consistency, average scores for adjectives in indefinite neuter DPs are approaching 90% for most of the older groups, suggests that most children have in fact acquired this rule. The locus of the problem therefore appears to be the failure to apply the rule, which is line with the MSIH. However, to test this proposal directly, non-production data on adjectives are needed.

8. Conclusion

This paper investigated the role of current and CUMULATIVE amount of exposure on the acquisition of

grammatical gender, as marked on definite determiners and adjectives in indefinite DPs, in English–Dutch simultaneous bilingual children. Current amount of exposure and CUMULATIVE length of exposure were both found to be significant predictors for gender-marking on determiners but not (directly) for gender-marking on adjectives. Using detailed parental questionnaire data allowed us to examine children's exposure patterns over time, in order to test the prediction that for bilingual children with relatively little exposure, acquisition may be “timed off the map” for certain target language properties and to investigate whether amount of exposure in the early years may play a role in subsequent language development. There was little evidence that this was the case for the target language property under investigation here. The finding that current amount of exposure was also a significant predictor variable underlines the importance of continued language exposure and use in the maintenance and success of bilingual acquisition, even for simultaneous bilingual children. Results from the grammaticality judgement task suggested that when children fail to produce the target definite determiner *het* with neuter nouns, this may result from a production-specific problem rather than having failed to specify the noun in question as [+neuter].

It is hoped that these findings will contribute to a growing body of research exploring the external and internal factors affecting bilingual language acquisition (input quantity/quality, socio-economic status, language use, etc. vs. age of onset, knowledge of another language, cognitive maturity, language learning aptitude, etc., respectively). It is only by systematically investigating a wide range of factors for different language combinations and linguistic properties that we can hope to arrive at a more complete understanding of how children acquiring more than one language can do so successfully.

Appendix

Table A1. Average percentage of nouns produced with target definite determiner.

	Monolingual children			Bilingual children		
	n	Common	Neuter	n	Common	Neuter
3 year olds				11	99.1% <i>SD</i> 3.0	3.5% <i>SD</i> 4.5
4 year olds	6	100% <i>SD</i> —	34.5% <i>SD</i> 24.6	13	96.7% <i>SD</i> 7.1	10.2% <i>SD</i> 19.7
5 year olds	11	99.5% <i>SD</i> 1.7	61.6% <i>SD</i> 35.0	14	99.6% <i>SD</i> 1.3	21.2% <i>SD</i> 28.8
6 year olds	9	90.3% <i>SD</i> 21.0	60.4% <i>SD</i> 34.4	15	96.5% <i>SD</i> 5.7	42.1% <i>SD</i> 37.1
7 year olds				14	98.1% <i>SD</i> 2.9	69.3% <i>SD</i> 35.2
8 year olds				15	97.7% <i>SD</i> 3.4	65.1% <i>SD</i> 44.0
9 year olds				8	98.6% <i>SD</i> 2.7	87.0% <i>SD</i> 21.2
10 year olds				11	98.7% <i>SD</i> 3.0	74.4% <i>SD</i> 36.1
11 year olds				13	96.8% <i>SD</i> 4.5	55.6% <i>SD</i> 44.0
12 and 13 year olds				11	98.0% <i>SD</i> 3.8	80.8% <i>SD</i> 32.3
14 to 17 year olds				11	96.6% <i>SD</i> 5.6	64.2% <i>SD</i> 32.9

Table A2. Average percentage of adjectives produced with target inflection

	Monolingual children			Bilingual children		
	n	Common	Neuter	n	Common	Neuter
3 year olds				11	99.2% <i>SD</i> 2.5	8.0% <i>SD</i> 8.8
4 year olds	6	98.6% <i>SD</i> 3.4	23.8% <i>SD</i> 20.6	13	93.6% <i>SD</i> 20.7	20.2% <i>SD</i> 29.6
5 year olds	11	99.2% <i>SD</i> 2.5	47.2% <i>SD</i> 28.5	14	96.6% <i>SD</i> 9.0	27.0% <i>SD</i> 27.0
6 year olds	9	98.1% <i>SD</i> 4.1	61.9% <i>SD</i> 28.3	15	88.5% <i>SD</i> 19.4	52.6% <i>SD</i> 32.3
7 year olds				14	98.0% <i>SD</i> 4.1	50.0% <i>SD</i> 36.3
8 year olds				15	94.8% <i>SD</i> 17.1	60.7% <i>SD</i> 40.3
9 year olds				8	100% <i>SD</i> —	81.1% <i>SD</i> 20.0
10 year olds				11	99.5% <i>SD</i> 1.7	80.3% <i>SD</i> 23.0
11 year olds				13	97.0% <i>SD</i> 4.9	58.6% <i>SD</i> 37.0
12 and 13 year olds				11	100% <i>SD</i> —	74.2% <i>SD</i> 35.5
14 to 17 year olds				11	100% <i>SD</i> —	52.5% <i>SD</i> 45.0

Table A3. Average percentage of nouns marked consistently with target determiner.

	Monolingual children			Bilingual children		
	n	Common	Neuter	n	Common	Neuter
3 year olds				11	98.5 <i>SD</i> 5.0	0 <i>SD</i> —
4 year olds	6	100 <i>SD</i> —	25 <i>SD</i> 31.2	13	97.4 <i>SD</i> 6.3	5.1 <i>SD</i> 18.5
5 year olds	11	100 <i>SD</i> —	60.0 <i>SD</i> 38.2	14	100 <i>SD</i> —	15.1 <i>SD</i> 27.0
6 year olds	9	89.2 <i>SD</i> 24.5	61.0 <i>SD</i> 40.9	15	96.9 <i>SD</i> 5.3	39.1 <i>SD</i> 37.1
7 year olds				14	98.4 <i>SD</i> 4.0	64.7 <i>SD</i> 38.1
8 year olds				15	99.2 <i>SD</i> 3.2	66.7 <i>SD</i> 48.8
9 year olds				8	100 <i>SD</i> —	89.4 <i>SD</i> 22.4
10 year olds				11	100 <i>SD</i> —	71.2 <i>SD</i> 40.0
11 year olds				13	97.4 <i>SD</i> 6.7	55.3 <i>SD</i> 47.3
12 and 13 year olds				11	99.0 <i>SD</i> 3.4	81.8 <i>SD</i> 32.3
14 to 17 year olds				11	98.0 <i>SD</i> 6.7	60.5 <i>SD</i> 37.5

Table A4. Average percentage of adjectives produced with target inflection, consistently-marked nouns only.

	Monolingual children				Bilingual children			
	Common		Neuter		Common		Neuter	
	n	%	n	%	n	%	n	%
3 year olds					11	97.7% <i>SD</i> 5.4	0	—
4 year olds	6	98.6% <i>SD</i> 3.4	4	41.7% <i>SD</i> 50.0	13	94.6% <i>SD</i> 19.6	1	100% <i>SD</i> —
5 year olds	11	99.2% <i>SD</i> 2.5	9	73.1% <i>SD</i> 21.2	14	96.4% <i>SD</i> 9.1	5	65.0% <i>SD</i> 41.8
6 year olds	9	98.8% <i>SD</i> 3.7	7	87.7% <i>SD</i> 16.3	15	91.5% <i>SD</i> 18.6	9	83.6% <i>SD</i> 29.2
7 year olds					14	98.8% <i>SD</i> 3.2	12	67.6% <i>SD</i> 36.8
8 year olds					15	95.2% <i>SD</i> 17.2	10	81.7% <i>SD</i> 27.7
9 year olds					8	100% <i>SD</i> —	8	87.8% <i>SD</i> 17.1
10 year olds					11	99.4% <i>SD</i> 1.8	10	84.4% <i>SD</i> 30.4
11 year olds					13	98.3% <i>SD</i> 4.2	8	92.4% <i>SD</i> 7.2
12 and 13 year olds					11	100% <i>SD</i> —	10	87.8% <i>SD</i> 18.5
14 to 17 year olds					11	100% <i>SD</i> —	10	68.3% <i>SD</i> 41.9

Table A5. Average percentage of target determiners selected in judgement task.

Group	n	Common	Neuter
5 year olds	12	77.8% <i>SD</i> 26.9	56.1% <i>SD</i> 28.4
6 year olds	15	85.6% <i>SD</i> 25.7	72.6% <i>SD</i> 26.8
7 year olds	14	92.1% <i>SD</i> 23.6	83.3% <i>SD</i> 20.3
8 year olds	15	96.3% <i>SD</i> 6.9	83.0% <i>SD</i> 24.8
9 year olds	8	100% <i>SD</i> —	93.1% <i>SD</i> 11.8
10 year olds	11	96.0% <i>SD</i> 10.3	91.9% <i>SD</i> 16.5
11 year olds	12	90.7% <i>SD</i> 15.6	83.3% <i>SD</i> 20.4
12 and 13 year olds	11	95.7% <i>SD</i> 10.3	90.9% <i>SD</i> 17.1
14 to 17 year olds	11	85.9% <i>SD</i> 21.1	85.9% <i>SD</i> 18.7

References

- Bianchi, G. (in press). Gender in Italian–German bilinguals: A comparison with German L2 learners of Italian. *Bilingualism: Language and Cognition*, doi:10.1017/S1366728911000745. Published by Cambridge University Press, February 10, 2012.
- Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *International Journal of Bilingualism*, 14, 422–446.
- Blom, E., Polišenská, D., & Weerman, F. (2008a). Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research*, 24, 297–332.
- Blom, E., Polišenská, D., & Unsworth, S. (2008b). The acquisition of grammatical gender in Dutch. *Second Language Research*, 24, 259–265.
- Blom, E., & Vasic, N. (2011). The production and processing of determiner–noun agreement in child L2 Dutch. *Linguistic Approaches to Bilingualism*, 1, 265–290.
- Brouwer, S., Cornips, L., & Hulk, A. (2008). Misrepresentation of Dutch neuter gender in older bilingual children? In E. Gavrusseva & B. Haznedar (eds.), *Trends in child second language acquisition*, pp. 83–96. Amsterdam: John Benjamins.
- Bruhn de Garavito, J., & White, L. (2002). The second language acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. Muñoz Liveras (eds.), *The acquisition of Spanish morphosyntax*, pp. 153–178. Amsterdam: John Benjamins.
- Carroll, S. E. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, 39, 535–594.
- Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry*, 31, 319–355.
- Chondrogiani, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, 1, 318–342.
- Cornips, L., & Hulk, A. (2006). External and internal factors in bilingual and bidialectal language development: Grammatical gender of the Dutch definite determiner. In

- C. Lefebvre, L. White & C. Jourdan (eds.), *L2 acquisition and creole genesis: Dialogues*, pp. 355–378. Amsterdam: John Benjamins.
- Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, 28, 267–296.
- Cornips, L., van der Hoek, M., & Verwer, R. (2006). The acquisition of grammatical gender in bilingual child acquisition of Dutch (by older Moroccan and Turkish children): The definite determiner, attributive adjective and relative pronoun. In B. Los & J. van de Weijer (eds.), *Linguistics in the Netherlands 2006*, pp. 40–51. Amsterdam: John Benjamins.
- Damhuis, R., de Gloppe, K., Boers, M., & Kienstra, M. (1992). *Woordenlijst voor 4- tot 6-jarigen. Een streeflijst voor kleuters*. Rotterdam: Projectbureau OVB.
- De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge: Cambridge University Press.
- De Houwer, A. (2009). *Bilingual first language acquisition*. Clevedon: Multilingual Matters.
- Deutsch, W., & Wijnen, F. (1985). The article's noun and the noun's article: Explorations into the representation and access of linguistic gender in Dutch. *Linguistics*, 23, 793–810.
- Donaldson, B. C. (1987). *Dutch reference grammar*. Leiden: Martinus Nijhoff.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (PPVT-4)* (4th edn.). Minneapolis, MN: Pearson.
- Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam: Pearson.
- Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale*. London: GL Assessment.
- Ellis, N. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27, 1–24.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Gathercole, V. C. M. (2002a). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In Oller & Eilers (eds.), pp. 207–219.
- Gathercole, V. C. M. (2002b). Command of the mass/count distinction in bilingual and monolingual children: An English morphosyntactic distinction. In Oller & Eilers (eds.), pp. 175–206.
- Gathercole, V. C. M. (2002c). Monolingual and bilingual acquisition: Learning different treatments of *that*-trace phenomena in English and Spanish. In Oller & Eilers (eds.), pp. 220–254.
- Gathercole, V. C. M., & Thomas, E. M. (2005). Minority language survival: Input factors influencing the acquisition of Welsh. In J. Cohen, K. McAlister, K. Rolstad & J. MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism*, pp. 852–874. Somerville, MA: Cascadilla Press.
- Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12, 213–237.
- Geerts, G., Haeseryn, W., Rooij, J. de, & Toorn, M. C. van de (1984). *Algemene Nederlandse spraakkunst*. Groningen: Wolters-Noordhoff.
- Genesee, F., & Nicoladis, E. (2007). Bilingual first language acquisition. In E. Hoff & P. McCordle (eds.), *Handbook of language development*, pp. 324–342. Oxford: Blackwell.
- Gilger, J. W. (1992). Using self-report and parental-report survey data to assess past and present academic achievement of adults and children. *Journal of Applied Developmental Psychology*, 13, 235–256.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28, 217–241.
- Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, 24, 267–288.
- Halle, M. (1997). Distributed morphology: Impoverishment and fission. In B. Bruening, Y. Kang & M. McGinnis (eds.), *Papers at the interface* (MIT Working Papers in Linguistics 30), pp. 425–449. Cambridge, MA: MIT Press.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (eds.), *The view from Building 20*, pp. 111–176. Cambridge, MA: MIT Press.
- Haznedar, B., & Schwartz, B. D. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes & A. Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, pp. 257–268. Somerville, MA: Cascadilla Press.
- Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, 24, 131–161.
- Kiparsky, P. (1973). “Elsewhere” in phonology. In S. Anderson & P. Kiparsky (eds.), *A Festschrift for Morris Halle*, pp. 93–106. New York: Holt, Rinehart and Winston.
- Kupisch, T., Müller, N., & Cantone, K. F. (2002). Gender in monolingual and bilingual first language acquisition. *Lingue e Linguaggio*, 1, 107–150.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research*, 24, 459–486.
- Meisel, J. M. (2007a). On autonomous syntactic development in multiple first language acquisition. In H. Caunt-Nulton, S. Kulatilake & I. Woo (eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development*, pp. 26–45. Somerville, MA: Cascadilla Press.
- Meisel, J. M. (2007b). Exploring the limits of the LAD. *Working Papers in Multilingualism*, 80, 3–31.
- Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, 28, 5–34.
- Montrul, S. A., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58, 503–553.
- Oller, D. K., & Eilers, R. E. (eds.). (2002). *Language and literacy in bilingual children*. Clevedon: Multilingual Matters.

- Paradis, J. (2010a). Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning*, 60, 651–680.
- Paradis, J. (2010b). Comparing typically-developing children and children with specific language impairment. In E. Blom & S. Unsworth (eds.), *Experimental methods in language acquisition research*, pp. 223–244. Amsterdam: John Benjamins.
- Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1, 213–237.
- Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or independent? *Studies in Second Language Acquisition*, 18, 1–25.
- Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18, 41–58.
- Polišenská, D. (2010). Dutch children's acquisition of verbal and adjectival inflection. Ph.D. dissertation, University of Amsterdam.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103–133.
- Roodenburg, J., & Hulk, A. (2008). Puzzles on grammatical gender. *Lingue e Linguaggio*, 7, 67–91.
- Schlyter, S., & Håkansson, G. (1994). Word order in Swedish as the first language, second language and weaker language in bilinguals. *Scandinavian Working Papers on Bilingualism*, 9, 49–66.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1–33.
- Stöhr, A., Akpınar, D., Bianchi, G., & Kupisch, T. (2012). Gender marking in Italian–German heritage speakers and L2-learners of German. In K. Braunmueller & C. Gabriel (eds.), *Multilingual individuals, multilingual societies (MIMS)*, pp. 153–170. Amsterdam: John Benjamins.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th edn.). Boston, MA: Allyn and Bacon.
- Thomas, E. M., & Gathercole, V. C. M. (2007). Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language*, 27, 251–278.
- Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24, 365–396.
- Unsworth, S. (in press). Assessing age of onset effects in (early) child L2. *Language Acquisition*.
- Unsworth, S., Argyri, E., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (in press). On the role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*.
- Unsworth, S., & Hulk, A. (2009). L1 acquisition of neuter gender in Dutch: Production and judgement. In J. Costa, A. Castro, M. Lobo & F. Pratas (eds.), *Proceedings of Generative Approaches to Language Acquisition 2009*, pp. 483–492. Cambridge: Cambridge Scholars Publishing.
- Van Berkum, J. J. A. (1996). The psycholinguistics of grammatical gender: Studies in language comprehension and production. Ph.D. dissertation, Max Planck Institute for Psycholinguistics.
- van der Velde, M. (2003). Déterminants et pronoms en néerlandais et en français: Syntaxe en acquisition. Ph.D. dissertation, Paris 8.
- Weerman, F., Duijnmeijer, I., & Orgassa, A. (2011). Effecten van SLI op nederlandse congruentie. *Nederlandse Taalkunde*, 16, 30–55.
- White, L., Valenzuela, E., Kozłowska-MacGregor, M., & Leung, Y. I. (2004). Gender and number agreement in non-native Spanish. *Applied Psycholinguistics*, 25, 105–133.