

L1 ACQUISITION OF NEUTER GENDER IN DUTCH: PRODUCTION AND JUDGEMENT

SHARON UNSWORTH & AAFKE HULK

1. Introduction

Previous research indicates that grammatical gender in Dutch, and in particular neuter gender, is typically acquired late, with children overgeneralising common gender forms of the determiner to neuter nouns until (at least) age 6 (e.g., Blom et al. 2008, Bol and Kuiken, 1988, Van der Velde, 2003). Almost all of these previous studies use production data only. It is possible that these data may underestimate children's knowledge of Dutch gender. In this paper, we present data from production *and* judgement tasks carried out with the same children in order to determine whether this is the case. Furthermore, we examine the relationship between production and judgement data to more fully explore the developmental trajectory involved in the acquisition of Dutch gender.

2. Acquisition of neuter gender in Dutch

Dutch has a two-way gender system that distinguishes between neuter and common nouns. Gender is marked overtly on definite, relative and demonstrative determiners and adjectives. The focus here is on definite determiners. Common nouns are preceded by the definite determiner “de”, as in “de deur” (“the door”), whereas neuter nouns combine with the definite determiner “het”, as in “het huis” (“the house”). The gender specification of a given noun is generally assumed to be arbitrary (e.g., Geerts et al. 1984, Donaldson, 1987). Limited evidence for the gender specification of certain nouns is available but the morphological and semantic regularities in question are few, they involve numerous exceptions, and perhaps more importantly, they very often involve abstract and/or complex nouns only, which young children are unlikely to be familiar with.¹ Essentially, the only evidence available to children for a

noun's gender specification comes from the few gender-bearing elements with which it appears. The evidence for neuter nouns is limited because (i) common nouns outnumber neuter nouns by about 2:1 (van Berkum, 1996), (ii) the singular definite determiner for common nouns "de" is used for plural nouns of both genders, and (iii) the neuter definite determiner "het" is often produced in reduced form which in certain phonological contexts makes it indistinguishable from "de". In sum, given the absence of clear-cut, age-appropriate semantic and morphological cues, we may expect that the acquisition of grammatical gender in Dutch takes quite some time.

This limited evidence for neuter gender in Dutch appears to contribute to the apparent delay which L1 children experience in its acquisition. Previous research indicates that monolingual L1 children overgeneralise "de" with neuter nouns until (at least) age 6 (e.g., Van der Velde 2003). Almost all of these previous studies are based on production data only. To our knowledge, there are only two studies on comprehension data. Johnson (2004) analyses preferential looking data from 28-month-old children, and finds that they are sensitive to gender-marking with "de" on common nouns but not to "het" on neuter nouns. Brouwer et al. (2008) analyse oral judgement data from 11- to 13-year-old children and find, perhaps unsurprisingly, that they perform at ceiling. The available "non-production" results to date are from children who either display no knowledge or completely targetlike knowledge of neuter gender.

In (adult) non-native (L2) acquisition, where grammatical gender often remains problematic, even for many learners who would otherwise be categorised as near-native, it has been suggested that there may be a discrepancy between learners' underlying morpho-syntactic representation and the realisation of the relevant morpho-syntactic forms in production (Lardiere, 2000). In other words, production data may underestimate a learner's knowledge of a target language property. More specifically, for gender, it has been suggested that while (L2) learners may have the target gender specification for certain nouns in their interlanguage grammars, they may experience problems in realising the correct form of gender-marked forms such as definite determiners in production (White et al. 2004, Unsworth 2008). If this is also the case for gender in L1 acquisition, we should observe differential performance on tasks involving production and comprehension. Whereas monolingual Dutch-speaking children may (incorrectly) make use of "de" with neuter nouns in production, in comprehension they should be (more) targetlike.²

In this paper, we examine production *and* judgement data from the same group of children who are old enough to produce some gender-marked forms but still young enough not to perform at ceiling. More

specifically, we investigate the hypothesis that 4- to 7-year-old children will be significantly better on a judgement than a production task. Note that this hypothesis is valid for neuter nouns only. Given that the definite determiner “de” is initially used with all nouns regardless of their gender, the (production) data of “de” will not necessarily tell us anything about the acquisition of common gender.

3. Participants

A total of 29 monolingual L1 children took part in this study (as part of a larger project). Given that the acquisition of Dutch gender to a certain extent involves word-learning, we decided to group children according to their scores on the standardised Peabody Picture Vocabulary Test (PPVT-III-NL), rather than according to their age, as this may provide us with a more accurate picture of how the development of gender proceeds. The scores, which are in general average or slightly above average, are presented in Table 1, along with the children’s ages.

Table 1. Overview of children’s ages and PPVT scores

Group	n	PPVT score			Age		
		Mean	SD	Range	Mean	SD	Range
I	15	101.3	5.3	95-108	6;0	1;1	4;4 – 7;6
II	14	117.8	7.3	109-129	5;7	0;8	4;3 – 6;8

Although the division into these two groups is essentially arbitrary such that two approximately equal-sized groups could be obtained, it is in some sense grounded in reality in that the children in group I have average vocabulary scores (falling within one SD of the average score of 100), whereas the children in group II are on average 1SD higher. The children who are in group II are thus at a (slightly) more advanced stage in their vocabulary development than the children in group I; we assume that vocabulary development can be taken as a general indicator for their overall linguistic development. Hence, when it comes to the acquisition of gender, we will also assume group II to be at a more advanced stage than group I.³

4. Production

4.1 Tasks

Two elicited production tasks were used: a picture description task (following Blom et al., 2008, a.o.) and a story task. In the first task, children are presented with two pictures on a computer screen, e.g., a yellow and a blue robot, and asked to name them using the following prompt designed to elicit adjectival agreement with indefinite nouns: “Look! Here we see two pictures. This is a ... (child: yellow robot). And this is a ... (child: blue robot)”. The part of the task which is relevant for the present paper is the following. An additional item, e.g., a ball, 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 twice with a definite determiner.⁴ Fillers were items testing verb form and placement (used for another part of the project).

A further, third token is elicited for each noun in the story task, where children help tell a story using pictures on the computer. Children are first asked to name the relevant nouns, and subsequently to name the same item in response to a question in order to elicit a definite NP. 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”.

Because these two tasks were part of a larger test battery, and because younger children have a shorter attention span, two versions of each task were used: one for younger children (≤ 5 years), where the maximum number of items per gender was 21 and one for older children (> 6 years), with a maximum of 27 items per gender.⁵ Each version was presented in two presentation orders, B being the reverse of A, counter-balanced across children. The results for the two tasks are presented together.

4.2 Results

There was no effect of presentation order for either common ($t(28) = -1.00, p > .05$) or neuter nouns ($t(28) = 1.44, p > .05$). The average number of items produced with a definite determiner by the younger children was 18.9 for common nouns and 18.3 for neuter nouns (maximum: 21),

and for older children, this was 24.9 for common nouns and 24.1 for neuter nouns (maximum: 27). Because the number of items varies per child, the results are presented as average percentages rather than raw numbers. The analysis presented here is the number of nouns produced with the target definite determiner (i.e., “de” or “het”) divided by the total number of nouns produced with a definite determiner (i.e., “de” plus “het”). The results are given in Table 2.

Table 2. Results for production: Average percentage of nouns in definite NP produced with target determiner

Group	Common	Neuter
I	95.4 <i>SD</i> 16.5	41.2 <i>SD</i> 33.7
II	98.3 <i>SD</i> 4.3	66.0 <i>SD</i> 30.3

Children produce significantly more target definite determiners with common nouns than with neuter nouns (group I: $t(14) = -4.88$, $p < .001$; group II: $t(13) = -3.80$, $p < .05$). There is a significant difference between groups for neuter nouns ($t(27) = -2.08$, $p < .05$) but not for common nouns ($t(27) = -.65$, $p = > .05$). There is considerably more within-group variation in children’s responses for neuter nouns than for common nouns

5. Judgement

5.1 Task

In order to assess children’s knowledge of gender using a “non-production” task, a grammaticality judgement task was developed where children have to make a forced choice concerning the acceptability of congruent vs. incongruent determiner-noun combinations. The nouns are identical to those used in the production tasks, and again, two versions were also used here. For reasons of time, each noun was tested once only, which meant a maximum number of 6 responses for the younger children (≤ 5 years) and 9 responses for the older children (> 6 years). The two presentation orders were counter-balanced across children.

Children are first presented with a set of five objects depicting three target nouns and two fillers. The experimenter first introduces each object

using an indefinite (“Here we see a ...”). Next, a close-up of one of these five objects is shown and two puppets, who are introduced as coming from a planet far away and as learners of Dutch, are asked to name what they see. Each puppet describes the object using a definite NP (now felicitous due to the previous introduction of the object by the experimenter with an indefinite NP). In a pre-recorded voice, one produces a target NP, e.g., “het huis” (“the house” where “huis” is a neuter noun), and the other its non-target equivalent, i.e., “de huis”. Target and non-target responses are counter-balanced across puppets and fillers are used to check that children are paying attention and for a puppet-bias. Fillers are of three types: (i) one puppet produces a nonsense noun, e.g., “de perg”, and the other the target noun corresponding to the picture, e.g., “de banaan”; (ii) one puppet produces the target form of a plural noun, e.g., “de poezen” (“the cats”) and the other reverses the order of the noun and determiner, as in “poezen de” (“cats the”); (iii) one puppet produces the target form of a plural noun, e.g., “de t-shirts”, and the other produces the same plural noun but with a non-target determiner, i.e., “het t-shirts”.

The child’s task is to decide which puppet is correct. Each puppet appears separately on the screen as it speaks and both are present on the screen when the child has to make a choice. They are told that one puppet is always right and so they must make a choice. In a pilot, we found that allowing children to choose neither or both puppets led to confusion, and for this reason, we opted for a forced choice task.

5.2 Results

There was no effect of presentation order for orders for common nouns ($t(28) = -1.05, p > .05$) or neuter nouns ($t(28) = -1.12, p > .05$). One child was excluded because of a puppet bias.⁶ The results for the remaining children are presented in Table 3. In order to facilitate the comparison with the production results, the data are presented as percentage correct here, too, i.e., the number of nouns for which the target determiner was selected divided by the total number of nouns with that gender.

Table 3. Results for judgement: Average percentage of nouns for which target definite determiner selected

Group	Common	Neuter
I	92.6 <i>SD</i> 7.5	61.9 <i>SD</i> 27.5
II	86.9 <i>SD</i> 18.3	77.0 <i>SD</i> 20.1

Children in group I are better at selecting the correct puppet with common nouns than with neuter nouns ($t(14) = -4.22$, $p = .01$), but there is no difference for group II ($t(13) = -1.20$, $p > .05$). There is no significant difference between group I and group II for either neuter ($t(27) = -1.66$, $p > .05$) or common ($t(27) = 1.11$, $p > .05$) nouns. As in the production data, there is also considerable within-group variation in the judgement data.

6. Production and judgement compared

When we compare the production and judgement data (including only those children for whom we have data for both modalities), we find that for common nouns, there is no significant difference between production and judgement data for group I ($t(14) = .620$, $p > .05$), but there is a difference for group II ($t(13) = 2.25$, $p < .05$). Group I perform equally well on both but group II perform significantly worse on judgement than production. For neuter nouns, there is a significant difference between production and judgement for group I ($t(14) = -2.63$, $p < .05$), but not for group II ($t(13) = -1.18$, $p > .05$). In other words, with neuter nouns, group I are significantly better on judgement than on production, whereas group II perform equally well, on both, albeit not at ceiling. The individual data for neuter nouns confirm these results: 25 of the 29 children perform better on judgement than on production or equally well on both.

7. Discussion

Let us first summarise the results. The results for production indicate that 4- to 7-year-old Dutch children overgeneralise the definite determiner “de” with neuter nouns. This is consistent with previous findings (e.g.,

Blom et al., 2008, Van der Velde 2003). For neuter nouns, there was a significant effect of group with children in the higher vocabulary score group producing significantly more neuter nouns with the target determiner “het”. It should be noted, however, that there is considerable variation between children for neuter nouns.

For the judgement task, the data show that children choose the puppet producing the congruent determiner-noun combination significantly more often for common nouns than for neuter nouns. In other words, they are more likely to falsely choose the non-target NP “de huis”, where there is a mismatch between a common gender determiner and a neuter gender noun, than the non-target NP “het telefoon”, where there is a mismatch between a neuter gender determiner and a common gender noun. In contrast to the production results, on the judgement task, there is no significant effect of group for either common or neuter nouns. For neuter nouns, this lack of a group effect seems to be due to the observation that group I are significantly better on judgement than on production. In other words, on the judgement task, the gap between the two groups has decreased. Again, there is considerable variation between children, especially for neuter nouns, but also for common nouns in group II.

When we compare the results for production and judgement, we find a significant difference between the two (i) for neuter nouns for group I and (ii) for common nouns for group II. This first result is in line with our hypothesis. The children’s superior performance on judgement may indicate that production tasks may underestimate their underlying knowledge of gender. It is important to note that the children in our sample fall within the range of the scores for production for comparably-aged children found in previous studies (e.g., van der Velde 2003). Note, however, that even though the group I children are significantly better on judgement than on production, their average judgement scores are still far from targetlike. The difference between production and judgement data may simply be a task effect. While this is of course a possibility, the observation that for group II, the observed significant difference between production and judgement on common nouns holds in the opposite direction, i.e., judgement is on average *worse* than production, indicates that any such task effect, if this exists, does not hold across the board.

What do these data tell us about the developmental path via which Dutch gender is acquired? The children in the present study are divided into two groups on the basis of their scores on a standardised vocabulary task. We assume that these scores may be taken as an indicator of general linguistic development and thus that the children in group I are at a less advanced stage of linguistic development than group II. The results for the

production of neuter nouns show a significant effect of group, which does indeed support the idea that the children in group II are more advanced than those in group I with respect to the production of “het”. It may be the case that the group II children are already aware that “de” and “het” are in a paradigmatic relation. What they now have to do is to specify for each noun whether it takes “het” or “de” as a definite determiner. At such a stage of development, we expect children to also show some hesitation with respect to common nouns, sometimes incorrectly assuming that certain common nouns are neuter. The significant difference between judgement and production with common nouns in group II may be seen as support for such a developmental step.⁷

In order to gain a complete picture of the relation between production and judgement and of how the development of gender in Dutch takes place, it would be necessary to examine data from a group of comparably aged children with lower vocabulary scores than group I here. The prediction would be that a group of children with lower vocabulary scores should be at a less advanced stage of linguistic development than group I and thus should perform more poorly on neuter nouns on the production tasks, and possibly also on the judgement task.

Notes

¹ The only exception is diminutives, which always take “het” and which are also frequent in the input to children. While there is evidence to suggest that monolingual children do indeed use this cue (Van der Velde 2003), we leave this type of noun outside the scope of this paper.

² The claim that comprehension may precede production is of course not new—see e.g., Clark and Hecht (1983). The data we report here are judgement data and hence can probably be more accurately characterised as “non-production” rather than “comprehension” in the strictest sense.

³ Note that in this particular sample of children, the younger children perform slightly better (for their age) than the older children.

⁴ The results for adjectival inflection (with definites as well as indefinites) are left aside here.

⁵ The selected nouns include words expected to appear at different frequencies in the input (based on a word list for 4- to 6-year-old children—Damhuis et al. 1992). An analysis of the results in terms of frequency is left for future research.

⁶ A puppet-bias was operationalised as follows: children who selected puppet A for (i) five consecutive items where puppet B was the target and (ii) all other items, i.e., where either puppet A or puppet B was target, in-between. The criteria of five consecutive items means that, due to the pseudo-randomised order in which items are presented, the child makes errors on all three types of items, i.e., fillers, de-words and het-words.

⁷ Blom et al. (2008) also mention some examples of incorrect overgeneralisation of ‘het’ to common nouns in L1 children around age 6, but not before (see also Unsworth 2008 for similar findings with L2 children and adults).

Acknowledgements

Thanks to Laura Tejedor Timomor for data collection and the Netherlands Organisation for Scientific Research (NWO) for funding.

References

- Blom, E., et al. (2008). “Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender.” *Second Language Research* 24, 297-331.
- Bol, G. & F. Kuiken. (1988). Grammaticale analyse van taalontwikkelingsstoornissen. PhD thesis, University of Amsterdam, The Netherlands.
- Brouwer, S. et al. (2008). Misrepresentation of Dutch neuter gender in older bilingual children? In *Trends in child second language acquisition*, eds. E. Gavrusseva & B. Haznedar, 83-96. Amsterdam: John Benjamins.
- Clark, E.V. & B.F. Hecht. (1983). “Comprehension, production and language acquisition.” *Annual Review of Psychology* 34, 325-349.
- Damhuis, R. et al. (1992). *Woordenlijst voor 4- tot 6-jarigen: Streeflijst voor kleuters*. Rotterdam: Projectbureau OVB.
- Donaldson, B.C. (1987). *Dutch reference grammar*. Leiden: Martinus Nijhoff.
- Geerts, G., et al. (1984). *Algemene nederlandse spraakkunst*. Groningen: Wolters-Noordhoff.
- Johnson, E. (2005). Grammatical gender and early word recognition in Dutch. In *BUCLD 28*, eds. Brugos et al., 320-330. Somerville, MA: Cascadilla Press.
- Lardiere, D. (2000). Mapping features to forms in second language acquisition. In *Second language acquisition and linguistic theory*, ed. J. Archibald, 102-129. Oxford: Blackwell.
- Unsworth, S. (2008). “Age and input in the acquisition of grammatical gender in Dutch.” *Second Language Acquisition* 24, 365-396.
- Van Berkum, J.J.A. (1996). The psycholinguistics of grammatical gender: Studies in language comprehension and production. PhD thesis, MPI Nijmegen.
- Van der Velde, M. (2003). Déterminants et pronoms en néerlandais et en français: Syntaxe en acquisition. PhD thesis, Paris 8.
- White, L., et al. (2004). “Gender and number agreement in nonnative Spanish.” *Applied Psycholinguistics* 25, 105-133.