Growing up with Frisian and Dutch: The role of language input in the early development of Frisian and Dutch among preschool children in Friesland
Dijkstra, J.E.

Citation for published version (APA):
Dijkstra, J. E. (2013). Growing up with Frisian and Dutch: The role of language input in the early development of Frisian and Dutch among preschool children in Friesland

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
3 Methodology

This chapter explains the methodology of the project: the design is presented (§ 3.1), followed by a description of the classification of the two independent variables based on the questionnaires on language input (§ 3.2). Furthermore, the non-verbal intelligence (§ 3.3) and the characteristics of the participants (§ 3.4) are described. Next, the assessment procedure, the instruments and the statistical analysis are explained (§ 3.5).

3.1 Research design

The aim of the project is to investigate the role of language input in the early development of a bilingual vocabulary in Frisian and Dutch. As already stated in § 2.6 and § 2.7, the participants were monitored in several aspects of language, i.e. receptive and productive vocabulary, number of different words (NDW) and mean length of utterance in words (MLUw). The design of the longitudinal project is partly pseudo-experimental and partly naturalistic, i.e. vocabulary tests were used to monitor the development in vocabulary and spontaneous speech samples were used to extract MLUw and information on lexical diversity (NDW). The research design is illustrated in Figure 3.1.

The participants were monitored during three successive test rounds. During each round, both languages were assessed. For each language, the test battery consisted of a receptive and a productive vocabulary test, followed by a recording of spontaneous speech while the participant was playing with the test assistant (see § 3.5). Data from the first round (age 2;6-3;0 years) functioned as a baseline measurement, which enabled comparison with the two follow-up rounds. Round 2 took place when the participants were aged between 3;0-3;6 years old and the last follow-up, round 3, was when the participants were aged between 3;6-4;0 years. During these follow-up assessments the participants were tested with the same test instruments except in round 3 when the participants were also tested on a non-verbal intelligence test (see § 3.3).
In sum, each participant was tested in seven test sessions within this project, i.e. in three Frisian and three Dutch sessions and a session in which non-verbal intelligence was tested. Consequently, at the end of the project we had gathered per child:
- three Frisian receptive vocabulary scores;
- three Dutch receptive vocabulary scores;
- three Frisian productive vocabulary scores;
- three Dutch productive vocabulary scores;
- three recordings of spontaneous speech in Frisian, yielding MLUw and NDW scores;
- three recordings of spontaneous speech in Dutch, yielding MLUw and NDW scores;
- information on input with questionnaires;
- one non-verbal intelligence score.

The questionnaires provided information on language input within and outside the home. These were used to classify all participants with respect to home language and outside home exposure (see § 3.2). The language measures were
analysed and controlled for non-verbal intelligence (see § 3.3). The results are presented in two chapters: Chapter 6 gives the results for the vocabulary measures and Chapter 7 presents the outcomes for the morphological measure and for the relationship between morphosyntax and vocabulary.

### 3.2 Questionnaires on language input

Parents as well as preschool teachers completed detailed questionnaires in the first test round. The questions were partly based on parental questionnaires used in a previous study (Dijkstra, 2008) and are for the most part comparable with Unsworth (2012). The questionnaires provided information on the child’s language input inside and outside the home or at preschool, and background information such as education level of the parents etc. The questions concerning language input inside and outside the home were repeated during every round to monitor input differences over time. The information from the parental questionnaires were used to identify the characteristics of the participants (see § 3.4.2 and § 3.4.3 for a description of the participants).

The parental questionnaires were used to classify the home language and outside home exposure of the participants. The home language is defined as the language that is predominantly used by both parents towards the participants. Both father and mother were asked to indicate which language(s) they spoke to their child: Frisian only, predominantly Frisian, 50% Frisian and 50% Dutch, predominantly Frisian or Frisian only. The home language was classified as Frisian (HL-Frisian) if both parents indicated they spoke only or predominantly Frisian to their children. Likewise, the home language was classified as Dutch (HL-Dutch) if both parents indicated they spoke only or predominantly Dutch to their children. Participants whose parents chose other combinations (e.g. 50% Frisian and 50% Dutch, and only Frisian, or predominantly Frisian and Dutch only) were excluded from the study, since the amount of input varied depending on the main caregiver. The questions on the language(s) that both father and mother used with their child was repeated in every test round. However, since the home language turned out to remain constant over time (round 1 x round 2: \( r = .91, p < .01 \); round 2 x round 3: \( r = .92, p < .01 \); round 1 x round 3: \( r = .92, p < .01 \), based on the large sample \( N = 91 \)), see § 3.4.2), it was decided to determine the home language variable per child to the classification of round 1: 58 HL-Frisian participants and 33 HL-Dutch participants.
The outside home exposure was defined as the language input from caregivers other than the parents, e.g. grandparents, preschool teachers, daycare providers, and private caregivers, during weekdays. With the parental questionnaires, information was gathered on the frequency (number of mornings and afternoons) that participants visited other caregivers during weekdays and the caregivers’ language use towards the participants. For example, parents had to indicate which mornings and/or afternoons the grandparents took care of their child (if so), and/or when the participant attended preschool or daycare. Next, they specified the languages used by these caregivers in communication with the participant. For each participant, the exposure pattern in Frisian and Dutch was calculated by adding the mornings and afternoons of outside home exposure to Frisian and to Dutch. If the caregiver’s exposure was 50% Frisian and 50% Dutch both language scores received a half a point, otherwise the language in question increased by one point. This resulted in two outcomes, one for Frisian and one for Dutch. Based on the ratio between these two outcomes, the participant’s outside home exposure was classified as exposure to the same language as the home language (OH-same language), or as outside home exposure to the other language (OH-other language).

The outcome of the language that was similar to the home language represented the category of outside home exposure to the same language as at home. The other outcome represented the outside home exposure to the other language. If the outcome of the outside home exposure to the other language was higher or roughly equal to that of the home language (at maximum the exposure to the other language was two mornings/afternoons less than the exposure to the home language), the participant was classified as being substantially exposed to the other language outside the home (OH-other language). Otherwise the outside home exposure was categorized as being similar to the home language (OH-same language). For example, an HL-Frisian participant had an outside home exposure of five mornings/afternoons to Frisian and three mornings/afternoons to Dutch. This means that the participant was substantially exposed to the other language (Dutch) and the outside home exposure of the participant was categorized as OH-other language. If the outside home exposure to Dutch had been two mornings/afternoons, the outside home exposure of the participant would have been categorized as OH-same language. The classification process for outside home exposure was repeated for the two follow-up rounds in order to monitor differences in outside home exposure. These differences are presented and further discussed in Table 3.1 in § 3.4.2 (large group, N=91), and in Table 3.2 in § 3.4.3 (subgroup, N=20).
Furthermore, background information was gathered on education, native language(s), proficiency in Frisian of the parents, participants’ health, family composition, number of children, etc. Moreover, data on the participants’ use of different media was collected via the questionnaires, e.g. watching television or DVDs, listening to radio or CDs and playing computer games (which claimed to stimulate language acquisition). Additionally, parents were asked how frequently and in which language(s) they read books to their children. The background information was used in the description of the participants’ characteristics (see § 3.4.2 and § 3.4.3).

The questionnaires of the preschool teachers were used to gather information on the preschool teachers’ language use in preschool during several activities. They additionally included questions about language behaviour of the participants in the playgroup. These questions were repeated during every round to monitor differences in input for the participants over time. The language(s) used by the preschool teacher in the playgroup and in personal contact with the child were used to identify the language input at preschool in the calculation of the outcomes for Frisian and Dutch within the classification of the participants’ outside home exposure variable as discussed above.

### 3.3 Non-verbal intelligence test

Several studies have revealed a relation between intelligence and performances on language tests (see § 2.1). Therefore, when interpreting the results of the language data, the intellectual performances of the participants must be taken into consideration. The non-verbal intelligence test, i.e. the SON-R 2.5-7 (Tellegen, Winkel, Wijnberg-Williams & Laros, 2005a), was used to measure intelligence in the current project, because this test was developed especially for young children within the age range 2.5-7 years. Another advantage of this test is that it is a non-verbal intelligence test. Consequently, the verbal instruction is kept to a minimum, which means that the test should elicit same results between children of different language backgrounds. Given the short concentration span of the children and the limited test time, it was not possible to administer the whole non-verbal intelligence test to the participants. Only two of the six subtests were selected for the current project. Since the ability to reason is an important characteristic of intelligence (Paradis, 2011; Carroll, 1993; Genesee & Hamayan, 1980), a subtest for abstract reasoning, Categorieën [Categories], and a subtest for concrete reasoning, Situaties [Situations], was used to
get an impression of the participants' intelligence. The reliability of both sub-
tests is relatively high in the ages 2;6, 3;6 and 4;6 years, i.e. respectively .81, .73, 
.70 for Categorieën, and .79, .66 and .62 for Situaties (Tellegen et al., 2005b: 53).
In the subtest Categorieën, the participants have to sort cards into two groups 
based on the pre-given category, or choose the card with the same features as 
the features on the three stimulus cards. In the subtest Situaties, the participants 
have to complete four drawings with complementary cards, or choose the cor-
rect card that shows the part that is missing on the picture. Both subtests have 
15 items and are aborted when the participant makes three mistakes in total.

Both subtests required minimal verbal instructions, and these were giv-
en in the participants' home language. The two test assistants who assessed all 
participants with the two subtests were both native speakers of Frisian and 
and had no previous history with the participants. In other words, they were com-
plete strangers to all participants. Before testing the participants, they were 
trained by a certified tester.

The non-verbal intelligence score in the current study was calculated by 
taking the sum of the correct responses (maximum score=30). This sum score 
representing the participant's non-verbal intelligence was used as a control 
variable in the analysis of the receptive and productive vocabulary scores.

3.4 Participants

This project has two sets of participants. All participants were assessed in recep-
tive and productive vocabulary. However, for the analyses of MLUw and NDW a 
subgroup of those participants was used. This subgroup is described separately. 
First, the selection method for participants is presented (§ 3.4.1). Then, the large 
group of participants from the receptive and productive vocabulary measures 
are described (§ 3.4.2), followed by the description of the subgroup (§ 3.4.3).

3.4.1 Selection of participants

Participants were selected by contacting preschool teachers of several preschools 
in the province of Friesland. If the preschool teacher was interested in participat-
ing in the project, the board of the preschool and sometimes the parental board 
also had to give their consent before children could be selected. If their permis-
sion was granted, the author contacted the preschool teacher again to select all 
children under the age of 2;10 from families where both parents spoke mainly 
Frisian or mainly Dutch to their children. Due to the age and the language
restriction, generally only one to three children from each preschool could be selected for this project. We asked the parents of these children for consent for their children to participate in the study. This procedure initially led to a sample of 101 participants. Such a large sample was necessary, because children this young of age are hard to test, so it was expected that not all participants would cooperate during the assessments. Three of these 101 participants came from mixed language families, which was discovered after parents had filled in their language questionnaires. These three participants were excluded from the study. Further, seven participants refused to carry out all vocabulary tests, moved away, or could not be tested in their non-verbal intelligence because they had already entered primary education. These participants were also excluded, ultimately leaving the final research group that comprised 91 participants.

In total, 23 preschools participated in the project, of which eleven preschools had a Frisian-medium or bilingual language policy, i.e. in these preschools both preschool teachers spoke Frisian only, or one teacher spoke Frisian only and the other Dutch only. Most preschools were located in rural parts of Friesland. Two preschools were located in small towns.

Since only a few participants per playgroup and/or preschool took part in the project, they all attended preschool on different days and times of day and the preschools were located in different regions of Friesland, it was difficult to schedule the assessments. A maximum of two participants could be tested during one morning. Moreover, the assessments could not be scheduled during the preschools’ holidays and celebrations, e.g. Sinterklaas (December 5th), Christmas and Easter. Since the process of selecting participants was very time-consuming and the first group of recruited participants soon approached the age of 3;0, i.e. the upper boundary of the first test round, the 91 participants were assessed in three cohorts. This means that the first group of participants started with the first assessments between September 2009 and January 2010, a second cohort of participants had their first assessments between February and July 2010, and the last cohort with participants started their first test round between September and October 2010.

For pragmatic reasons a subgroup of 20 participants was selected for the analysis of the spontaneous speech samples. First, a pre-selection was made from all participants with six language samples, i.e. three in Frisian and three in Dutch, that lasted longer than eleven minutes so that each sample would generate at least 100 child utterances. From that group, 20 participants were selected on home language (Frisian vs. Dutch) and gender. The two home language groups were matched on non-verbal intelligence.
3.4.2 Large group (N=91)

The vocabulary study comprised 91 participants. Based on the questionnaires the participants were classified into HL-Dutch and HL-Frisian (see also §3.2). The majority, i.e. 58, had parents who both spoke Frisian to them (HL-Frisian participants). The other 33 participants had parents who both spoke Dutch to them (HL-Dutch participants). Table 3.1 gives an overview of the characteristics of the participants.

<table>
<thead>
<tr>
<th></th>
<th>HL-Frisian</th>
<th>HL-Dutch</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boy</td>
<td>33</td>
<td>15</td>
<td>48</td>
</tr>
<tr>
<td>Girl</td>
<td>25</td>
<td>18</td>
<td>43</td>
</tr>
<tr>
<td>SES(^a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>24</td>
<td>18</td>
<td>42</td>
</tr>
<tr>
<td>Low</td>
<td>34</td>
<td>15</td>
<td>49</td>
</tr>
<tr>
<td>Intelligence(^b)</td>
<td>Mean (SD)</td>
<td>13.24 (3.03)</td>
<td>13.67 (3.16)</td>
</tr>
<tr>
<td></td>
<td>Min-Max (Max=30)</td>
<td>7-21</td>
<td>7-18</td>
</tr>
<tr>
<td>Age at start</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>2;9 (0;1)</td>
<td>2;8 (0;1)</td>
<td>2;9 (0;1)</td>
</tr>
<tr>
<td>Range</td>
<td>2;6-2;11</td>
<td>2;6-2;11</td>
<td>2;6-2;11</td>
</tr>
<tr>
<td>Outside home</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same as HL(^c)</td>
<td>43</td>
<td>19</td>
<td>62</td>
</tr>
<tr>
<td>Round 1 Other language</td>
<td>15</td>
<td>14</td>
<td>29</td>
</tr>
<tr>
<td>Round 2 Same as HL(^c)</td>
<td>41</td>
<td>14</td>
<td>55</td>
</tr>
<tr>
<td>Other language</td>
<td>17</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>Round 3 Same as HL(^c)</td>
<td>40</td>
<td>17</td>
<td>57</td>
</tr>
<tr>
<td>Other language</td>
<td>18</td>
<td>16</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>58</td>
<td>33</td>
<td>91</td>
</tr>
</tbody>
</table>

Table 3.1: The characteristics of the participants per home language group.

\(^a\) SES = socio-economic status
\(^b\) Subtests Situaties [Situations] and Categorieën [Categories] of non-verbal intelligence test SON-R 2,5-7 (Tellegen et al., 2005a)
\(^c\) HL = home language

The socio-economic status (SES) was based on parental education. The SES was rated high when one or both parents had a HAVO diploma or higher, i.e. a degree of higher general secondary education or pre-university education. No differences were found between the HL-Frisian participants and the HL-Dutch participants in gender, SES, or non-verbal intelligence (resp. \(\chi^2(1)=1.11, p>.05\); \(\chi^2(1)=1.47, p>.05\); \(t(89)=-.63, p>.05\)).
Table 3.1 displays the outside home exposure of the participants during every round. As Table 3.1 shows, the outside home exposure differed mostly between rounds 1 and 2, and generally remained constant between rounds 2 and 3 (round 1 x round 2: $\rho=0.60$, $p<0.01$; round 2 x round 3: $\rho=0.91$, $p<0.01$; round 1 x round 3 $\rho=0.69$, $p<0.01$). Furthermore, there was no dependency between participants’ home language and their exposure to the other language outside the home, except in round 2 (round 1: $\chi^2(1)=2.66$, $p>0.05$; round 2: $\chi^2(1)=7.03$, $p<0.01$; round 3: $\chi^2(1)=2.74$, $p>0.05$). This suggested that, in general, the language(s) used in the caregiver network of the parents showed no significant relation with the parents’ own main language use at home. Conversely, this relation was significant in round 2, which might be explained by a shift in the number of HL-Dutch participants who had a substantial exposure to the other language (Frisian) in that round. This shift might be the result of social desirable responses from the parents, since in round 3 the number of HL-Dutch participants who were exposed to Frisian outside the home decreases again (see Table 3.1).

All 91 participants were assessed in the main vocabulary study, i.e. in receptive and productive vocabulary. However, because of their young age they were hard to test. For example, they did not cooperate, or were not present at the preschool when the assessments took place, etc. The missing data points in the main vocabulary study are described in § 6.1.

The information described from here onward till the end of this subsection functions as background information on the participants. The information was derived from the parental questionnaires of the first test round.

Overall, the participants lived in various rural areas in Friesland. Five participants lived in small towns. During the entire period of data collection, two participants were referred to a speech and language therapist because of a suspected delay in speech and/or language development. However, they were not removed from the project, since their test scores did not deviate substantially from the other participants. All other participants had a normal language and general development according to the information from the parental questionnaires.

The families had on average 2.23 children. The group of participants consisted of 42 (46%) first-borns. The average number of older siblings was 0.80 (range 0–5). During the first test round, almost all parents in the HL-Dutch group reported their participants spoke only Dutch during solo play. Two (6%) HL-Dutch participants sometimes used Frisian as well. These two participants had an OH-exposure to the other language. Only one of these two
children had siblings and sometimes used Frisian when playing with them. In contrast with the HL-Dutch group, the HL-Frisian participants used the other language more often than the one spoken at home during play. In the HL-Frisian group, 31 (53%) participants sometimes used Dutch as well as Frisian during solo play. Further, 25 (43%) HL-Frisian participants sometimes used Dutch during play with their sibling(s). The same trend was seen for the two follow-up test rounds, which means that over time the language behaviour during play did not change substantially.

The parental questionnaires revealed that the HL-Frisian participants were regularly read to in both Frisian and Dutch. Dutch books were used slightly more often during reading than Frisian in this home language group, i.e. 78% of the HL-Frisian participants were read to in Dutch more than once a week and 74% were read to in Frisian more than once a week. All HL-Dutch participants were read to in Dutch more than once a week. Their parents read far fewer Frisian books to them. About 36% of the HL-Dutch participants were never read to in Frisian. Interestingly, an equally large part of this group, 36%, was sometimes read to in Frisian whereas 21% (i.e. seven participants) was even regularly read to in Frisian. This finding is a bit surprising since one would not expect the Dutch-speaking parents to read books in Frisian to their children and the answer to this question might be influenced by social desirability.

Additionally, all participants regularly listened to Dutch radio or music CDs, or watched Dutch television or DVDs. Frisian media were used to a much smaller degree, and mostly by HL-Frisian participants. Of course, when interpreting these results one has to consider the substantially smaller availability of Frisian media compared to Dutch, e.g. Dutch children’s programmes are available on several television channels, whereas for Frisian there is only one channel with one children’s programme per day. In conclusion, the parental questionnaires revealed that taking book reading and media into consideration, the HL-Frisian participants generally received more input in Dutch, compared to the HL-Dutch participants’ input in Frisian.

From the age of 3;0 years onwards, all participants attended preschool for two to three mornings/afternoons per week. In the case of 38 participants (43%), the preschool was a certified Frisian-medium preschool or bilingual preschool where the preschool teachers spoke Frisian only, or one teacher spoke Frisian only while the other used Dutch only. The other 53 participants (58%) attended a preschool without a language policy, i.e. at these preschools the teachers spoke mostly Dutch and sometimes Frisian, for example in individual contact with the children, provided the teacher was able to speak Frisian.
3.4.3 Subgroup (N=20)

As mentioned in § 3.4.1, it was only possible to transcribe the samples of 20 of the 91 participants due to time limitations. Only those participants were selected who had six samples of more than eleven minutes of spontaneous speech, resulting in at least 100 child utterances per sample. From this pre-selection 20 participants were selected and matched on home language, gender and non-verbal intelligence. Ten participants had thus Frisian as their home language (HL-Frisian participants) and ten Dutch (HL-Dutch participants). Each home language group consisted of five boys and five girls who were matched on non-verbal intelligence with the participants in the other home language group. Table 3.2 gives an overview of the characteristics of the participants.

<table>
<thead>
<tr>
<th></th>
<th>HL-Frisian</th>
<th>HL-Dutch</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boy</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Girl</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>SES(^a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>7</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Intelligence(^b)</td>
<td>Mean (SD)</td>
<td>Min-Max (Max=30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.00 (3.02)</td>
<td>9-18</td>
<td>14.10 (3.01)</td>
</tr>
<tr>
<td></td>
<td>14.20 (3.16)</td>
<td>9-18</td>
<td></td>
</tr>
<tr>
<td>Age at start</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) Range</td>
<td>2;9 (0;1)</td>
<td>2;6-2;11</td>
<td>2;8 (0;1)</td>
</tr>
<tr>
<td></td>
<td>2;7 (0;1)</td>
<td>2;6-2;10</td>
<td></td>
</tr>
<tr>
<td>Outside home</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 1 Same as HL(^c)</td>
<td>7</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Other language</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Round 2 Same as HL(^c)</td>
<td>6</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Other language</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Round 3 Same as HL(^c)</td>
<td>6</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Other language</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 3.2: The characteristics of the subgroup per home language group.

\(^a\) SES = socio-economic status

\(^b\) Subtests Situaties [Situations] and Categorieën [Categories] of non-verbal intelligence test SON-R 2,5-7 (Tellegen et al., 2005a)

\(^c\) HL = home language
As described in § 3.2, outside home exposure was measured in each of the three successive rounds of measurements. The outcomes are displayed in Table 3.2. The outside home exposure during the third round was most stable with equal outside home exposure for HL-Frisian participants in round 2 and similar outside home exposure for HL-Dutch participants in round 1, these outcomes were used in the analyses of MLUw and NDW. One of the conditions in the selection process of this subgroup was that all participants had three Frisian samples and three Dutch samples that lasted over eleven minutes each, so that each sample would generate the 100 child utterances needed for the analyses (see § 3.4.1). As a consequence, the language measures MLUw and NDW have no missing data points.

The following information was derived from the parental questionnaires of the first test round and used as background information. All participants in the subgroup were read to in Dutch and except for one HL-Frisian participant this happened at least once per week. All HL-Frisian participants were read to in Frisian. In contrast, six HL-Dutch participants were never read to in Frisian, whereas this happened at least sometimes for the remaining four HL-Dutch participants. All participants regularly used Dutch media, while Frisian media were used far less and only by HL-Frisian participants. The participating families lived in various rural areas of Friesland and had on average 2.30 children. The participants all had a normal language and general development. The group of participants consisted of eight (40%) first-borns. The average number of older siblings for HL-Frisian participants was 1.40 (range 0-2). Six HL-Frisian participants had two older siblings. In contrast, six HL-Dutch participants were firstborns; they had on average 0.50 older siblings (range 0-2). The parents of HL-Dutch participants reported that their children spoke Dutch only with their sibling(s), while the HL-Frisian participants mostly used Frisian with their sibling(s). However, six HL-Frisian participants sometimes spoke Dutch as well, for example during play. Five (25%), i.e. three HL-Frisian participants and two HL-Dutch participants attended a Frisian-medium or bilingual preschool. The other 15 participants went to preschools that did not have an explicit language policy. Dutch was the main language used in these preschools.
3.5 Assessment procedure, instruments and analysis

This section first explains the assessment procedure (§ 3.5.1). It continues with the language measures and instruments used in the project: receptive vocabulary tests (§ 3.5.2), and productive vocabulary test (§ 3.5.3), NDW (§ 3.5.4) and MLUw (§ 3.5.5). The section ends with a description of the statistical analyses (§ 3.5.6).

3.5.1 Assessment procedure

Six test assistants conducted the assessments. The test assistants were instructed to follow the one person-one language approach (Döpke, 1992). Hence, the test assistants who assessed the participants in Frisian spoke Frisian only, no matter what language the participant used. Likewise, the test assistants who assessed the participants in Dutch spoke Dutch only. In this way, the target language was elicited as much as possible.

The assistants were trained during two sessions. During the first session, the assistants became acquainted with the vocabulary tests and the test procedure. In the following two weeks they practised the test procedure with young children (not participating in the project) and during the second session their experiences were evaluated. All six test assistants were female and lived in Friesland. The three test assistants for Frisian were all native or near-native speakers of Frisian. Two of the three assistants for Dutch were native speakers of Dutch and had difficulties understanding Frisian. The third assistant for Dutch was a native speaker of Frisian.

With some exceptions, all assessments took place at preschool during the morning sessions. During the assessments, it was of great importance that participants felt at ease. Therefore, the test assistant first gained their confidence by playing with them and their friends in the playgroup before taking them individually outside the classroom for the assessment. Furthermore, it was essential to take the short attention span of these young participants into consideration. When the participants showed signs of fatigue, the test assistant included a break by bringing them back to the playgroup and finishing the assessment later that morning. The order of the vocabulary tests also played a role in making participants feel at ease. The test assistants started with the receptive vocabulary test, which only required the participants to point to pictures. Next, they were tested with the productive vocabulary test followed by the recording of spontaneous speech. By that time all participants were generally used to the test situation and felt free to talk. The spontaneous speech samples
were gathered at each round by recording a conversation between the test assistant and the participant while they were playing with a standard set of toys (O’Brien & Nagle, 1987), i.e. miniature animals and a puzzle with large pieces that represent rooms of a house. Used to generate lexical diversity (NDW) and morphosyntactic measure, i.e. mean length of utterance. See § 3.5.4 and § 3.5.5 for the transcription procedure and the generation of both measures.

The research design consisted of a series of repeated measurements, so two follow-up rounds came after the baseline measurements in the first round (see also § 3.1). During every round, both languages were tested on separate occasions with a few weeks in between. Therefore, when aged 2;6-3;0 years, the participants were first tested in Frisian and a few weeks later in Dutch. This procedure was repeated when participants were aged 3;0-3;6 years, and again when they were 3;6-4;0 years old. Moreover, it should be noted that the Frisian and the Dutch assessments in test rounds 1 and 2 could not be exactly spaced six months apart, because of the language switch between the test rounds, and because the assessments had to be scheduled around holidays and celebrations (see § 3.4.1). This means that the interval between the Frisian and Dutch baseline or follow-up measures was three to eight months. In the case of a three-months interval a child was, for example, tested at age 2;10 (round 1) in Dutch and again at 3;1 years (round 2). In the case of an eight-months interval a child was, for example, assessed in Frisian at age 2;7 (round 1) and again at age 3;3 (round 2).

Unfortunately, not all participants could be tested when they were aged between 2;6 and 3;0 years old in the baseline situation. In some cases it took a long time to get the parental consent forms back. Consequently, some participants were (just) over 3;0 years old when they were tested for the first time. In other cases, participants were slightly older than 3;0 years in the first measurements because at that time it was uncertain if it was possible to get enough participants and every participant was more than welcome. Due to their older age they were only tested twice and their data were included in rounds 2 and 3. To avoid language order biasing the results, the order of the languages was switched during the second round, i.e. participants were assessed in Dutch first and in Frisian next. For the third round, the language order was Frisian-Dutch again. Ideally, the language assessments should be counterbalanced, with half of the participants being assessed in the Frisian-Dutch order and the other half in the Dutch-Frisian order. However, for organizational reasons, e.g. the large number of participants, the distance between locations of the preschools, the fact that participants all attended preschool during different
daily periods, the fact that no more than two participants could be tested in a single morning session, and the fact that holidays and other celebrations had to be avoided, but also for reasons of time planning considerations and limited financial resources, counterbalancing was not an option in the current project.

3.5.2 Receptive vocabulary test
This section describes the receptive vocabulary tests used in the main vocabulary study. It starts with the Dutch receptive vocabulary test (§ 3.5.2.1) and continues with the Frisian vocabulary test (§ 3.5.2.2).

3.5.2.1 Dutch
Dutch receptive vocabulary was assessed with the Peabody Picture Vocabulary Test-III-NL (PPVT-III-NL) (Schlichting, 2005a). This is a Dutch adaptation of the American-English Peabody Picture Vocabulary Test-III (Dunn & Dunn, 1997). The Dutch Peabody Picture Vocabulary Test uses a set of four pictorial stimuli, arranged in a 2x2 frame on a page. The participant indicates which of the four pictures matches the stimulus word; the other three pictures are distractors. This vocabulary test has 17 sets comprising a total of 204 items. Each set contains twelve items that are ordered according to difficulty. This vocabulary test is standardized for the ages 2;3-90 years. For this study, only the 108 items of the first nine sets were selected, because we were dealing with very young children and items 109-204 were assumed to exceed their maximum ability range. The internal consistency of this Dutch receptive vocabulary test is high: the Guttman reliability score (Guttman, 1945) is $\lambda = .91$ in the age group 2;6-2;11 years (Schlichting, 2005b: 45). According to the standard procedure, the vocabulary test was aborted when the participant made nine or more mistakes in a twelve-item set. The total score on this vocabulary test was the sum of all correct items with a maximum score of 108 points. Because we wanted to assess trends in vocabulary growth, raw scores were used in preference to standardized scores.

3.5.2.2 Frisian
There was no test available that assessed the receptive vocabulary of Frisian. The Dutch vocabulary test Peabody Picture Vocabulary Test-III-NL was therefore adapted to Frisian. Chapter 4 describes the adaptation process (see § 4.1) and the pilot that was conducted with this adaptation (see § 4.3). The abortion procedure and the calculation of the total score were similar to the Dutch receptive vocabulary test. The maximum score was 108 points.
3.5.3 Productive vocabulary test

This section describes the productive vocabulary tests used in the main vocabulary study. It starts with the Dutch productive vocabulary test (§ 3.5.3.1), followed by the Frisian vocabulary test (§ 3.5.3.2).

3.5.3.1 Dutch

Dutch productive vocabulary was assessed with the Subtest Woordontwikkeling [Word development] of the Dutch Schlichting Test voor Taalproductie II [Schlichting Test for Language Production II] (Schlichting & Lutje Spelberg, 2010a). In this vocabulary test, the participant is shown one picture and is asked to finish a stimulus sentence, mostly by naming the picture. The productive vocabulary test consists of 70 items that are ranked in order of difficulty. The internal consistency of the subtest Woordontwikkeling is high (λ-2= .88, age 2;9 years) (Schlichting & Lutje Spelberg, 2010b: 28).

Input in one language might be restricted to certain contexts, e.g. the home, while the other language is mainly used in other situations, e.g. at preschool. Therefore, bilingual children often develop a different lexicon in both languages. In other words, it might occur that some words are only known in one language and not in the other. This phenomenon is also called a ‘distributed characteristic’ (Oller et al., 2007; Oller, 2005; Oller & Pearson, 2002) (see also § 2.3). When testing bilingual children, some test items might elicit an incorrect response while they would have elicited the correct response in the other language. The vocabulary study took the distributed characteristic into account by adapting the standard abortion procedure: correct responses in the language other than the language tested did not play a role in the abortion procedure of the productive vocabulary tests. For example, when the Frisian item glydzje ‘to slide’ elicited an incorrect Dutch response, e.g. glijden ‘to slide’, in the Frisian assessments, it was not included in the abortion procedure. Only incorrect responses in both languages were included in this procedure, e.g. lopen ‘to walk’ in the example of glydzje. Correct responses included correct cognate responses (words that have the same pronunciation and meaning in both languages), Frisian specific responses, or Dutch specific responses (see § 5.3 for a more detailed definition of cognates). For example, the test item ‘truck’ had frachtauto and vrachtauto (same pronunciation and meaning in respectively Frisian and Dutch) as correct cognate response, its synonym frachtwein as correct Frisian specific response and vrachtwagen as correct Dutch specific response. A mixed response, e.g. vrachtwagen-tsje ‘(small) truck’ Dutch equivalent for truck with Frisian diminutive marker -tsje, was scored as an incorrect response and, consequently,
left out of the total score calculations. In this way, if an expected effect was absent in an analysis, this would not be caused by these mixed responses.

In addition, the total score only included the correct responses of the language tested, so for the Dutch productive vocabulary test the total score was the sum of the correct cognate responses and the correct Dutch specific responses, with a maximum score of 70 points.

3.5.3.2 Frisian
There was no Frisian productive vocabulary test available. Therefore, the Dutch subtest Woordontwikkeling was adapted to Frisian. The adaptation process and the pilot that was conducted with this adaptation are described in § 4.2 and § 4.3 respectively. The abortion procedure was similar to the Dutch productive vocabulary test. The total score was the sum of correct specific Frisian and cognate responses, with a maximum score of 69 points.

3.5.4 Number of different words (NDW)
As described in § 3.5.1, spontaneous speech samples were gathered to generate information on the lexical diversity in the child’s speech, i.e. number of different words, and also on the morphosyntactic measure used in this project, i.e. mean length of utterance (see § 3.5.5). This section will explain how these speech samples were transcribed and how the number of different words were generated from the child utterances in those samples.

Four research assistants who were not present during the recordings transcribed the child’s speech and the speech of the test assistants that played with the child in the recordings (see also § 3.5.1). These four research assistants were all native speakers of Frisian (and fluent in Dutch). They transcribed the first 100 child utterances per language sample, i.e. 600 utterances per participant in total. The transcriptions were made in the program PRAAT (Boersma & Weenink, 2012). An utterance was defined as ‘the minimal conversational contribution, consisting of one or more clauses’ (Wells, 1985: 60). Additional information such as pauses and intonation was also used to determine the boundaries of an utterance (Wells, 1985). Ungrammaticalities in the child utterances, such as incorrect inflections, were transcribed literally. The transcriptions therefore also contained ungrammatical structures or utterances. The author checked all transcriptions with the recordings afterwards, in order to verify that all child utterances were literally transcribed. Furthermore, all transcriptions were checked for inconsistencies in spelling and language classification (see below) in order to prevent wrong analyses.
In the transcription, every word in the child’s utterances was classified as a Frisian specific word, as a Dutch specific word, as a mixture of the two (e.g. paard-sje ‘(small) horse’ Dutch equivalent for horse with Frisian diminutive marker -sje), or as a cognate word. A cognate word has the same pronunciation and meaning in both languages (see § 5.3 for a more detailed definition of cognates). Words were classified as Frisian specific words based on the online version of the dictionary Wurdboek fan de Fryske Taal [Dictionary of the Frisian Language] (WFT) (Fryske Akademy, 2010), a large dictionary with information on pronunciation, regional varieties and etymological background. New words that had not yet been included in the WFT dictionary were collected on a special list so that if they occurred in later speech samples, they would be treated consistently. The Dutch equivalent, the online Woordenboek der Nederlandse Taal [Dictionary of the Dutch Language] (WNT) (Instituut voor Nederlandse Lexicologie/Nederlandse Taalunie, 2010) was used to classify words as Dutch specific. For coding and analysing, a special self-developed codebook based on the CHAT conventions of the program CLAN1 was used.

CLAN generated information on lexical diversity, i.e. the number of different words (NDW). The measure NDW consisted of the sum of all different (or unique) words used in the child utterances. For NDW only intelligible utterances were used, excluding repetitions within the child’s utterances, for example ik ik ik pak deze ‘I I I take this one’ has three unique words (ik pak deze). Furthermore, social expressions, e.g. utterances such as hoi ‘hi’ or kijk eens ‘look!’, and yes/no-utterances were excluded. Onomatopoeias such as, woef ‘woof’, fillers, such as eh, hm, oh, ah, and unintelligible words were also excluded from the NDW calculations. Because of time limitations, it was not possible to conduct a morphological analysis on the child data. CLAN therefore generated NDW by treating a word stem and morphological inflections with that stem as separate different words. This means that forms such as hûn ‘dog’ and hûn-tsje ‘(small) dog’ were treated as two different words, instead of one, i.e. the lemma hûn. The NDW analyses in the present study therefore reflect both the number of different words including the inflected forms.

For each round, the measures were computed for the Frisian sessions and for the Dutch sessions. For Frisian, only the Frisian specific words and cognates were used, excluding the Dutch loanwords, i.e. Dutch specific words.

---

1 CLAN stands for Computerized Language ANalysis. It is part of CHILDES, or CHild Language Data Exchange System, a transcription and analysis program (MacWhinney, 2000).
Likewise, we only used the Dutch specific words and cognates, excluding the Frisian loanwords, for the Dutch sessions. The words that were mixtures of Frisian and Dutch, e.g. *paard-sje* ‘(small) horse’ Dutch equivalent for horse with the Frisian diminutive marker -*jé*, were also left out of the analyses, because we wanted to make sure that if an expected effect was absent in an analysis, this would not be caused by the presence of mixed words.

### 3.5.5 Mean length of utterance in words (MLUw)

As described in § 3.5.1, the spontaneous speech samples were also used to collect data on the morphosyntactic growth, i.e. mean length of utterance in words (MLUw). The transcription procedure is explained in § 3.5.4. Mean length of utterance in words (MLUw) is widely used as a rough indicator of morphosyntactic complexity of child language (Bowers & Vasilyeva, 2011; Parker & Brorson, 2005). Using mean length of utterance as a measure may cause difficulties when morphosyntactic growth is compared across languages, especially when it concerns two different morphological types of languages, for example Dutch and Turkish. Turkish differs from Dutch in that it is a highly agglutinative language, so numerous affixes may be attached to a word stem. Consequently, mean length of utterance measured in words (MLUw) will not be comparable: the Turkish sample will generate a much lower MLUw than the Dutch sample. In such cases measures of mean length of utterance in morphemes (MLUm) as proposed by Brown (1973) is preferred over MLUw. A morpheme is the smallest meaningful unit of grammatical form, e.g. the root of a word, affixes and suffixes (Crystal et al., 1989). MLUm also has its disadvantages: sometimes ambiguous decisions have to be made when calculating the number of morphemes, since the guidelines of Brown (1973) cannot easily be applied to other languages (Parker & Brorson, 2005; Thordardottir & Weismer, 1998; Hickey, 1991; Arlman-Rupp, van Niekerk-de Haan & van de Sandt-Koenderman, 1976).

Within languages, MLUms and MLUw appear to correlate highly: not only in Dutch (Arlman-Rupp et al., 1976), but also in English (Parker & Brorson, 2005), Icelandic (Thordardottir & Weismer, 1998) and Irish (Hickey, 1991). Arlman-Rupp et al. (1976) therefore prefer the use of MLUw over MLUms since the measurement of MLUw is less time-consuming, easier to apply and more reliable. Since MLUms and MLUw are highly correlated within two other West Germanic languages, i.e. Dutch (Arlman-Rupp et al., 1976) and English (Parker & Brorson, 2005), it is expected that the same will be true for MLUms and MLUw within Frisian. In line with Arlman-Rupp et al. (1976), MLUw...
was therefore preferable to MLUm as a measure of morphosyntactic growth of Frisian and Dutch within the morphosyntactic study. As stated in § 2.2, the morphosyntactic structures in Frisian and Dutch child language are quite similar, when the Frisian version of LARSP (Dijkstra, 2008) is compared with one of the Dutch versions (Schlichting, 2005c). There is one Frisian morphosyntactic structure that might slightly influence the MLUw scores, i.e. the Frisian pronoun for second person singular, do ‘you (sg.)’ (see also Popkema, 2006; Tiersma, 1999). This pronoun allows subject pronoun drop, i.e. it can be omitted, as in example (1):

(1) Ø yt-st in appel.
    (you) eat-PRS.2SG in.ART.INDF apple
    ‘you are eating an apple.’

Or it can occur as a clitic directly after the verb, as in example (2):

(2) yt-st-o in appel.
    eat-PRS.2SG-2SG in.ART.INDF apple
    ‘you are eating an apple.’

Additionally, it can occur as a clitic directly after a subordinating conjunction, as in example (3):

(3) ik sjoch dat-st(o) in appel yt-st.
    1SG see-PRS.1SG that-2SG in.ART.INDF apple eat-PRS.2SG
    ‘I see that you are eating an apple’.

In contrast, Dutch always requires an explicit subject and this subject never occurs as a clitic. However, it is generally expected that this morphosyntactic structure will barely influence the total MLUw measure. We therefore argue that in the morphosyntactic study MLUw can be used to monitor and compare the morphosyntactic growth in both Frisian and Dutch.

MLUw was generated with CLAN (see also § 3.5.4) and is calculated by taking all words produced by the participant in a sample and dividing them by the number of utterances in that sample. As for the NDW (§ 3.5.4), the MLUw computation excluded unintelligible utterances, yes/no-utterances, and social expressions such as hoi ‘hi’, kijk eens ‘look!’. Fillers such as eh, hm, oh, ah, onomatopoeias, such as woef ‘woof’, and repetitions within the utterances, were
excluded from the utterances. For example, an utterance as ik ik ik pak deze ‘I I I take this one’ was treated as a three-word-utterance, i.e. ik pak deze ‘ik pak deze’.

For each round, the measures were computed for the Frisian sessions and for the Dutch sessions. For Frisian, the Frisian specific words and cognates were used, excluding the Dutch loanwords, i.e. Dutch specific words. Likewise, we only used the Dutch specific words and cognates, excluding the Frisian loanwords, for the Dutch sessions. As with NDW (see § 3.5.4), words that were mixtures of both languages, e.g. paard-sje ‘(small) horse’ Dutch equivalent for horse with Frisian diminutive marker -sje, were also excluded from the analyses, because we wanted to make sure that if an expected effect was absent in an analysis, this would not be caused by the presence of mixed words.

3.5.6 Statistical analysis
Data consisted of six vocabulary scores and one morphosyntactic measure. These data were analysed differently, because the number of participants in the two studies differed significantly. The main vocabulary study had 91 participants, while the data on MLUw and NDW were generated from a subgroup of 20 participants.

The receptive and productive vocabulary scores were analysed using multilevel modelling (SPSS, 2005), since the data was not equidistant, nor were participants tested at exactly the same ages, which is assumed when using repeated measures ANOVA. The advantage of multilevel modelling is not only that it increases statistical power, but also that it accounts for missing data (Quené & van den Bergh, 2008, 2004), i.e. if one participant missed one assessment the data from the other two test rounds would still be included in the results. This is important, since the participants in the current project were very young and did not always cooperate during every assessment. The missing data was both at random and by design, since four participants were too old during round 1, so they were only tested twice. Their data were included in round 2 and round 3. The data of a few outliers were also excluded from the results (see Table 6.1 in § 6.1 for a description of these missing data and an overview of the number of missing data points per vocabulary test). Raw scores were chosen over standardized scores because we wanted to assess trends in vocabulary growth, without monolingual norm referencing by chronological age. Each vocabulary test was analysed in a separate multilevel model, i.e. in total four final models were constructed. Each analysis started with an unconditional model (M0) without any of the variables. This unconditional model (M0) was controlled for substantial variance in intercepts.
across participants, i.e. the intercepts across participants were treated as random effects. One by one, variables were added to that model to see if the fit of the new model was significantly improved compared to the previous model. Further, the model was checked for possible interactions between variables. The final model (M1) shows the parameter estimates for the significant variables and interactions for the vocabulary test under investigation. As with the unconditional models (M0), the final models (M1) were also controlled for substantial variance in intercepts across participants. The results are presented in § 6.2 and § 6.3.

The data from the subgroup of participants was not suitable for multi-level modeling, since the relatively small number of observations and participants did not yield enough power. Instead these data were analysed using repeated measures ANOVA to test the effect of time and group (home language input and outside home language input) on the expressive measures of NDW (see § 6.4) and MLUw (see § 7.2). Correlation analysis was used to examine the relation between the vocabulary measures (see § 6.4) and between the measures MLUw, NDW and the receptive and productive vocabulary scores of the 20 participants from the subgroup (see § 7.4).