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
5 The cognate issue in the measurement of a bilingual vocabulary\textsuperscript{6,7}

This chapter deals with the issue of the presence of cognate items in vocabulary tests, especially in the case when comparing proficiency in two languages. To what extent do these cognate items influence the results of a language test? First, the issue is introduced by addressing several problems that lead to the research question (§ 5.1). In the next section (§ 5.2), previous research on the use of cognate items in vocabulary tests is presented. Then, the definition of a cognate in the present study is given (§ 5.3), followed by a description of the methodology (§ 5.4). The results of the receptive and productive vocabulary tests will verify the presence of the cognates (§ 5.5). Finally, the results are discussed (§ 5.6).

5.1 Introduction

Language proficiency in young children is often measured by means of a vocabulary test, because vocabulary is one of the keystones of language development. However, sometimes no vocabulary test is available for the language in question. In those circumstances, an existing vocabulary test is often adapted from another language into the target language. This is also the case in a large-scale project on the early bilingual language development of preschool-aged children living in the bilingual province of Friesland, in the north of the Netherlands. In this project, the bilingual participants were assessed with respect to their receptive and productive vocabulary in both Dutch, the dominant and national language, and Frisian, the minority language spoken in that region. However, no vocabulary tests were available for Frisian. Therefore, two standardized Dutch vocabulary tests were adapted to Frisian: a Dutch receptive


\textsuperscript{7} A slightly adapted version of this chapter has been submitted for publication as: Dijkstra, J., Klinkenberg, E. L., Kuiken, F., & Jorna, R. J. (submitted). Using cognates in vocabulary tests in language-proficiency assessment of bilingual speakers of closely related languages.
vocabulary test, i.e. the Dutch version of the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997; Dutch version Schlichting, 2005), and a Dutch productive vocabulary test, i.e. the subtest Woordontwikkeling [Productive Vocabulary] of the Dutch Schlichting Test voor Taalproductie II (Schlichting & Lutje Spelberg, 2010a) (see Chapter 4 for a description of the vocabulary tests, the adaptation process and the pilot that was conducted with these two vocabulary tests). The two languages in question have a fair amount of words in common because they are both West Germanic languages and because of a long history of language contact. This overlap in vocabulary was reflected in the two Frisian vocabulary tests, since they contain a considerable amount of cognate items, i.e. words that have the same pronunciation and meaning in both languages. The question then arises how well such Frisian test adaptations assess children’s Frisian vocabulary if they contain cognate items. In other words, is a vocabulary test containing cognate items still a valid test in the assessment of vocabulary knowledge in bilingual children’s two languages? Since the Dutch vocabulary tests also contain cognate items for native speakers of Frisian, both original Dutch vocabulary tests may be compromised in their measurement of Dutch vocabulary of these Frisian-speaking children. This validity issue leads to the following methodological question: do cognate items in the Frisian vocabulary tests and the Dutch vocabulary tests affect performance on the tests? The present study, carried out in the project just mentioned, was designed to resolve this issue. The results discussed in this study concern the outcomes of the baseline measurements using both the Frisian as well as the Dutch vocabulary tests.

5.2 Previous research on the use of cognates in test adaptations

Only a few studies have been conducted that focus specifically on the influence of cognates in vocabulary tests. Meara, Lightbown and Halter (1994) hypothesized that an assessment of a test-taker’s L2 with a vocabulary test that has a substantial amount of cognates in the test-taker’s L1 could lead to an overestimation of that test-taker’s overall proficiency. Conversely, the exclusion of those cognates from a vocabulary test in a language where these words naturally occur, could lead to an underestimation of the test-taker’s vocabulary size in that language. The researchers tested 89 native speakers of French living in Quebec with two versions of an English vocabulary test. Half of all items of one version were cognates, whereas the other version had none. As
The cognate issue in the measurement of a bilingual vocabulary

expected, the participants achieved higher scores on the cognate version compared to the non-cognate version. Although the difference in scores between both versions was significant, this difference was rather small. Meara et al. (1994) argued that their cognate version contained an unusually high number of cognates, substantially higher than the actual number of cognates between French and English. If their cognate version had consisted of a lower number of cognates, closer to the actual proportion of shared cognates between both languages, they would have expected to find no substantial difference in the scores between both versions. Therefore, Meara et al. (1994) claimed that as long as the number of cognates in a vocabulary test remains close to the actual number of cognates shared by the two languages, the presence of cognates will not significantly influence the outcomes of that vocabulary test. Their claim was in line with the findings of a study conducted by Umbel, Pearson, Fernández and Oller (1992), who compared the receptive vocabulary in Spanish and English of Mexican-American children. The cognate items in the Spanish and English vocabulary test did not play a substantial role in the performance of these bilingual children. Moreover, Umbel et al. (1992) showed that in bilingual language acquisition, cognates were not ‘automatic doublets’ for the children. McNamara and Roever (2006) further pointed out that cognates might make items easier, but they will not affect items’ relevance. Therefore, the cognate item remains a legitimate item in a vocabulary test. Knowledge of the first language will give participants an advantage, but only in the same way and to the same extent as in real life situations (McNamara & Roever, 2006; Chen & Henning, 1985).

Nevertheless, some researchers are against the use of cognate items in vocabulary tests. Gathercole et al. (2008) advised that cognate items should be avoided, even if they are not borrowings. Their Welsh receptive vocabulary test thus does not contain cognate items or English loanwords. Although these words might be used every day and have become part of everyday Welsh usage, Gathercole et al. (2008) argue that test-takers might respond appropriately to a cognate item based solely on their knowledge of English. In other words, it is impossible to reconstruct whether a cognate item assesses the test-taker’s English or Welsh proficiency.

Here Meara et al. (1994) is followed, that is vocabulary tests may contain cognate items, as long as the number of cognate items is in proportion to the actual number of cognates shared by the two languages. However, it is difficult to determine the exact number of cognates between two languages. This would require intense corpus work. It can be assumed that the proportion of
cognates will be larger in two languages that are closely related, for example for Frisian and Dutch. Since the exact proportion of cognates between Frisian and Dutch is not known, this proportion must be approached as an estimation. Nevertheless, Frisian and Dutch have a considerable overlap in vocabulary and this overlap should be reflected in the vocabulary tests of the two languages.

5.3 Cognates

As stated in § 5.1, Frisian and Dutch are related languages and have many similarities in vocabulary. The vocabulary tests in the two languages have thus a fair amount of cognate items in common. A cognate is often defined as a word that has identical meaning, spelling and pronunciation, e.g. *auto* in both languages (Frisian and Dutch equivalent for ‘car’). Since children in the age range of 2;6-4;0 years in the Netherlands have not developed reading and writing language skills yet, we defined cognates as being words with the same meaning and pronunciation only, excluding spelling. Words such as *fiets* (Dutch for ‘bike’) and *fyts* (Frisian for ‘bike’) are therefore considered cognates, since they are pronounced the same way. It is commonly assumed that morphological inflected cognates are also considered as cognates, even if these inflected forms exhibit slight between-language differences, e.g. *fiets-je* (Dutch for ‘(small) bike’) and *fyts-ke* (Frisian for ‘(small) bike’). In our study we considered words with the same stem followed by a language-specific marker as non-cognates, thus *fietsje* was considered as a Dutch-specific word and *fytske* as a Frisian-specific word. Especially when assessing productive vocabulary, these language-specific markers identify a response as either Frisian specific or Dutch specific. Another example is the word *appel* ‘apple’. The singular form is treated as a Frisian/Dutch cognate. This also applies to the plural *appels* ‘apples’, since the plural marker is identical in both languages. However in the diminutive the forms are different: *appeltje* ‘small apple’ (with the Dutch diminutive suffix -tje) vs. *appeltsje* ‘small apple’ (with the Frisian diminutive marker -tsje).

As already mentioned in § 3.5.2, the receptive vocabulary test contained 108 items in total. The Frisian adaptation had 46 (43%) cognate items compared to 47 (44%) cognate items in the Dutch version. The percentages differ slightly because, some items were replaced in the adapted version (see also § 4.1). In the productive vocabulary test there were 216 possible correct responses spread over 70 items. There were 160 correct language-specific
The cognate issue in the measurement of a bilingual vocabulary (non-cognates) responses, i.e. 88 Frisian specific and 72 Dutch specific responses. Of the 216 correct responses, only 26% (56) were cognate responses. This is much lower compared to the receptive vocabulary test, due to fact that items in this test can have more than one correct response. For example, the item ‘truck’ can elicit the cognate responses frachtauto (Frisian) and vrachtauto (Dutch), and the language-specific responses frachtwein (Frisian) and vracht-wagen (Dutch).

5.4 Methodology

The research group consisted of 97 participants, 49 boys and 48 girls, with a mean age of 2.9 years (SD 0.2, minimum 2.4, maximum 3.4) at the time of the baseline measurements. A large part of this group, i.e. 61 participants, grew up in families where both parents mainly spoke Frisian to them, whereas 36 participants grew up in families where Dutch was the home language. The participants were tested individually in a separate room at their preschool. First they were tested with the Frisian receptive vocabulary test, followed by the Frisian productive vocabulary test. A few weeks later this procedure was repeated with the Dutch vocabulary tests (see § 3.5.2 and § 3.5.3 for a description of these vocabulary tests). During the assessments the test assistants used the one person-one language principle (Döpke, 1992), i.e. the test assistants for Frisian only spoke Frisian with the participants and the test assistants for Dutch only spoke Dutch (see § 3.5.1 for more information on the assessment procedure).

The receptive vocabulary tests were aborted according to the standard procedure when the child produced nine or more incorrect responses in a twelve-item set. The total score on the receptive vocabulary tests was the sum of all correct items. The productive vocabulary test was aborted when the participant made eight errors in succession. The total score on the Frisian productive vocabulary test was calculated as the sum of all correct cognate responses and all correct Frisian specific responses, and similarly for Dutch.

8 This number is higher then the 91 participants mentioned in § 3.4.2 since it includes six participants who were excluded from the main vocabulary study because they could not be tested on non-verbal intelligence.
5.5 Results

This section presents the total sum scores of the first measurements of the Frisian and Dutch receptive vocabulary test (§ 5.5.1) and of the Frisian and Dutch productive vocabulary test (§ 5.5.2). We used raw scores only. For every vocabulary test, we calculated three scores, i.e. total correct score, the correct non-cognate (NC) items or responses, and cognate (C) items or responses. The participants are divided into two groups: those with Frisian as their home language (HL-Frisian participants) and those with Dutch as their home language (HL-Dutch participants). The mean per home language group will be compared to the overall mean of the three scores. If the total score of a vocabulary test differs significantly between the two home language groups, this significant difference should also be reflected in the NC scores. If the NC scores do not show this significant difference while there is a significant difference between both home language groups on the total score, the C items or responses must have influenced the results.

To get a better understanding of the data in the tables presented in the next subsections, it is worth mentioning again that we used abortion procedures during the assessments (see § 5.4). Although in theory participants could get a maximum total score, i.e. 108 points on the receptive vocabulary tests and 70 points (Dutch) or 69 points (Frisian) on the productive vocabulary tests, this maximum score was almost never reached, since practically all participants stranded somewhere in the middle of the test.

5.5.1 Receptive vocabulary tests

Table 5.1 shows the participants’ raw scores on the Frisian receptive vocabulary test grouped per home language, i.e. HL-Frisian participants and HL-Dutch participants, and the mean and standard deviation of the total group. First, the total score is given, followed by the performance on the NC items of the vocabulary test and the performance on the C items. Since this is a very difficult age group to assess, we only succeeded in testing 91 of the 97 participants with this vocabulary test. There were several reasons for this, e.g. anxiety, refusal or absence of the participants.

---

9 Since multiple responses per item were correct in the productive vocabulary tests, we use the term ‘response’ instead of ‘item’ for the productive vocabulary tests.
The cognate issue in the measurement of a bilingual vocabulary

Table 5.1: Frisian receptive vocabulary test: means (M) and standard deviations (SD) of the total scores of all items, the non-cognate item set (NC items), and the cognate item set (C items).

<table>
<thead>
<tr>
<th>Home Language</th>
<th>Total score</th>
<th>NC items</th>
<th>C items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M'</td>
<td>SD</td>
</tr>
<tr>
<td>Frisian</td>
<td>56</td>
<td>34.95</td>
<td>9.71</td>
</tr>
<tr>
<td>Dutch</td>
<td>35</td>
<td>29.63</td>
<td>7.88</td>
</tr>
<tr>
<td>Total</td>
<td>91</td>
<td>32.90</td>
<td>9.38</td>
</tr>
</tbody>
</table>

Note: Maximum total score is 108 points; Δtot denotes the difference between the mean of a home language group and the mean of all participants for each item set

* p<.01, ** p<.001

Table 5.1 presents the mean scores and standard deviations of the total score of all items, the NC items and the C items. The differences between the mean of the home language groups and the overall mean are listed under Δtot. For the whole vocabulary test, this was (34.95-32.90=) 2.05 points for the HL-Frisian group and (29.63-32.90=) -3.27 points for the HL-Dutch group. For the NC items the Δtot was 1.87 points for the HL-Frisian group and -2.99 for the HL-Dutch group. In contrast, the C items had a Δtot in both home language groups that was almost zero (see Table 5.1). The low Δtot for the C items implies that the cognate items did not cause the difference in scores between the two home language groups. Furthermore, the mean scores of the entire vocabulary test showed a significant difference between the two home language groups, (t(89)=2.73, p<.01, r=.28). This difference between both groups was reflected in a significant difference in mean scores on the NC items (t(89)=4.24, p<.001, r=.41). Conversely, no difference was found between the two groups in the C items (t(89)=.48, p>.05, r=.05). These results suggest that the performance of both home language groups on the NC items was in line with their performance on the vocabulary test as a whole. The presence of the C items did therefore not affect performance on the Frisian receptive vocabulary test.

Of course, the Dutch vocabulary tests also contained items that were equal in pronunciation and meaning in both languages. Therefore, it was also important to look at the results of the Dutch vocabulary tests. Table 5.2 presents the mean scores and standard deviations for the Dutch receptive vocabulary test.
Table 5.2: Dutch receptive vocabulary test: means (M) and standard deviations (SD) of the total scores of all items, the non-cognate item set (NC items), and the cognate item set (C items).

<table>
<thead>
<tr>
<th>Home Language</th>
<th>N</th>
<th>M* (SD)</th>
<th>Δtot</th>
<th>M* (SD)</th>
<th>Δtot</th>
<th>M (SD)</th>
<th>Δtot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frisian</td>
<td>61</td>
<td>36.61 (8.61)</td>
<td>-0.65</td>
<td>20.49 (4.81)</td>
<td>-0.22</td>
<td>16.11 (4.34)</td>
<td>-0.44</td>
</tr>
<tr>
<td>Dutch</td>
<td>31</td>
<td>38.55 (7.94)</td>
<td>1.29</td>
<td>21.13 (3.98)</td>
<td>0.42</td>
<td>17.42 (4.28)</td>
<td>0.87</td>
</tr>
<tr>
<td>Total</td>
<td>92</td>
<td>37.26 (8.40)</td>
<td></td>
<td>20.71 (4.53)</td>
<td></td>
<td>16.55 (4.34)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Maximum total score is 108 points; Δtot denotes the difference between the mean of a home language group and the mean of all participants for each item set.

The numbers (N) differ from the ones in Table 5.1, because not all participants could be successfully assessed with this vocabulary test. As Table 5.2 shows, both home language groups had comparable mean scores on both the NC items and on the C items. This is also reflected in the differences between the mean per home language group and the overall mean for the total score of this vocabulary test (see Δtot under Total score in Table 5.2), i.e. -0.65 points for the HL-Frisian group and 1.29 for the HL-Dutch group. The differences can, for the most part, be explained by the Δtot values on the C items, which were slightly larger than the ones on the NC items, i.e. -0.44 points and 0.87 for the HL-Frisian and HL-Dutch group on the C items, compared to -0.22 points and 0.42 respectively on the NC items. For the Dutch receptive vocabulary test, the scores on the C items did not differ significantly between the home language groups (t(90)=1.37, p>.05, r=.14), nor was there a significant difference between both home language groups on the NC items (t(90)=-0.64, p>.05, r=.07). Regarding the items of the entire vocabulary test, there the two home language groups did not differ either (t(90)=-1.05, p>.05, r=.11). Both HL-Dutch and HL-Frisian participants seemed to be equally competent in the receptive vocabulary of Dutch. Based on these outcomes, it is not possible to conclude whether the cognate items in this vocabulary test did or did not affect the test performance.

### 5.5.2 Productive vocabulary tests

Table 5.3 shows the mean scores and standard deviations on the productive vocabulary test for Frisian, separately for all items, the NC responses and the C responses. The difference in number of participants (N) between this
The cognate issue in the measurement of a bilingual vocabulary test and both receptive vocabulary tests can be explained by the fact that not every participant could be tested successfully with this vocabulary test.

<table>
<thead>
<tr>
<th>Home Language</th>
<th>Total score</th>
<th>NC responses</th>
<th>C responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M*</td>
<td>SD</td>
</tr>
<tr>
<td>Frisian</td>
<td>57</td>
<td>10.47</td>
<td>4.01</td>
</tr>
<tr>
<td>Dutch</td>
<td>34</td>
<td>6.59</td>
<td>2.46</td>
</tr>
<tr>
<td>Total</td>
<td>91</td>
<td>9.02</td>
<td>3.97</td>
</tr>
</tbody>
</table>

Table 5.3: Frisian productive vocabulary test: means (M) and standard deviations (SD) of the total scores of all items, the non-cognate responses (NC responses), and the cognate responses (C responses).

Note: Maximum total score is 70 points; Δtot denotes the difference between the mean of a home language group and the mean of all participants for each item set.

As Table 5.3 shows, the HL-Dutch participants had a mean score of almost zero (i.e. 0.35 points) on the NC responses, whereas on the C responses they had a mean score of 6.26 points, which was slightly higher than their HL-Frisian peers on this item set. These results show that the HL-Dutch participants mostly used a (correct) cognate word as a response on this vocabulary test. They gave scarcely any Frisian-specific response.

The differences between the group mean and the overall mean (Δtot) on the NC responses were 1.83 points for the HL-Frisian group and -3.08 for the HL-Dutch group (see Table 5.3), whereas they were almost zero for both groups on the C responses. This, again, indicates that the NC responses determined the total scores for the most part. For the Frisian productive vocabulary test there were significant differences found in scores between both home language groups both in the entire vocabulary test ($t(88.89)=5.10$, $p<.001$, $r=.48$), as well as in the NC responses ($t(66.96)=13.57$, $p<.001$, $r=.86$), in contrast to the C responses where they were not significant ($t(89)=-.70$, $p>.05$, $r=.07$). The HL-Dutch participants obtained the same results as the HL-Frisian participants on the C responses; however, they got lower results on the NC responses, which was in line with their results on the entire vocabulary test. Again, with this vocabulary test the C responses did not affect the results significantly.
The results of the Dutch productive vocabulary test are presented in Table 5.4. The number of participants (N) differs from those in the other three vocabulary tests, since we did not manage to test all participants successfully with this vocabulary test.

<table>
<thead>
<tr>
<th>Home Language</th>
<th>Total score</th>
<th>NC responses</th>
<th>C responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  M** SD Δtot</td>
<td>M** SD Δtot</td>
<td>M  SD Δtot</td>
</tr>
<tr>
<td>Frisian</td>
<td>57 8.65 4.51 -2.28</td>
<td>2.07 2.24 -2.17</td>
<td>6.75 2.92 -0.32</td>
</tr>
<tr>
<td>Dutch</td>
<td>31 15.13 6.56 4.20</td>
<td>8.23 4.36 3.99</td>
<td>7.65 3.02 0.58</td>
</tr>
</tbody>
</table>

Table 5.4: Dutch productive vocabulary test: means (M) and standard deviations (SD) of the total scores of all items, the non-cognate responses (NC responses), and the cognate responses (C responses).

Note: Maximum total score is 70 points; Δtot denotes the difference between the mean of a home language group and the mean of all participants for each set of responses * p<.01, ** p<.001

As Table 5.4 shows, the difference in mean between the home language group and the overall mean on the NC responses is -2.17 points for the HL-Frisian participants and 3.99 points for the HL-Dutch group. These differences were very small on the C responses, namely -0.32 points for the HL-Frisian group and 0.58 points for the HL-Dutch group. This implies that the total scores on this vocabulary test were mostly defined by the NC responses, and not by the C responses. Moreover, in the NC responses, the difference between the two home language groups was significant (t(38.84)=-7.36, p<.001, r=.76), in contrast to the C responses (t(86)=-1.35, p>.05, r=.14). The total scores were also significantly different (t(45.68)=-4.90, p<.001, r=.59). The HL-Dutch participants performed significantly better in the NC responses, whereas on the C responses their scores were equal to their HL-Frisian peers. With respect to the total score, the HL-Dutch participants scored significantly higher than their HL-Frisian peers. Consequently, it can be concluded that the C responses did not affect the results for the productive vocabulary test either.
5.6 Discussion and conclusions

The present study investigated whether the presence of cognate items affected the performance on vocabulary tests in bilingual proficiency assessments of closely related languages, i.e. Frisian and Dutch. The results showed that the cognate items did not have a significant influence on either the Frisian receptive vocabulary test or the productive vocabulary tests in both languages. On these vocabulary tests the participants from both home language groups performed equally well on the cognate items, whereas their scores differed substantially on the non-cognate items. The results on the Dutch receptive vocabulary test were unclear, since there were no significant differences in performance on this vocabulary test between the two home language groups. Differences between two groups become less important as their differences in competence become smaller, which was the case for this vocabulary test. A possible explanation is that both home language groups received input in Dutch, in contrast to Frisian where this was not always the case. It is commonly assumed that a higher input in a language will result in a higher proficiency in that language. Consequently, the proficiency in Dutch receptive vocabulary showed smaller differences in both home language groups, compared to Frisian receptive vocabulary. Another possible explanation is the fact that Dutch was tested a few weeks later than Frisian. By that time, the participants were used to a test situation outside the playgroup with a stranger and felt at ease when taking the vocabulary test. However, the results on the Dutch receptive vocabulary test were not mirrored in the results of the Dutch productive vocabulary test, which makes the last explanation less plausible.

As the results of the other three vocabulary tests in the present study suggest, the variance in test scores can not be explained by the presence of cognate items. The outcomes of the current study confirm the results of the previous study by Umbel et al. (1992) and the claim of Meara et al. (1994) that, as long as the number of cognate items reflects the actual proportion of cognates shared between two languages, the presence of the cognates does not jeopardize the validity of the vocabulary test. It should be recognized, however, that it is hard to determine the actual number of cognates shared between two languages. This requires intense corpus work. For most languages the proportion of shared cognates will always be an estimation.

Cognates are a natural part of a language. The items of vocabulary tests should represent the structure of the vocabulary tested. This is also true for a language that shows an overlap in vocabulary with another language.
Therefore, that overlap should then be present in the items of the vocabulary test. It is possible to be cautious and to delete the cognate items in retrospect. If the cognates are left out of the vocabulary test beforehand, the test will be more difficult for the test-taker. Moreover, such a cognate-free vocabulary test will focus on a higher level of competence instead of measuring the general level of vocabulary. Therefore, as long as the proportion of cognates in a vocabulary test reflects the proportion of cognates shared between two languages, we agree with Meara et al. (1994) and conclude that the presence of cognates in vocabulary tests is acceptable.