The acquisition of reference: a cross-linguistic study
Rozendaal, M.I.

Citation for published version (APA):

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.
4 Methodology

4.1 Introduction
This study examines the interaction between morphosyntax and pragmatics in the acquisition of reference with nouns, pronouns and proper names in children learning Dutch, English or French. The role of the input and cross-linguistic differences in acquisition are also studied (see the research questions in §3.5). Data from spontaneous conversation offer the best possibility of investigating these issues. The children studied are between two and three years-of-age and children of this age are usually too young to engage in experiments or narrations (Thornton, 1996; Wigglesworth, 1997). Moreover, longitudinal data enable the investigation of the development of reference. Finally, data are needed on the input that children receive. The selection of the subjects and data are described in §4.2 and §4.3, followed by a description of the analysis procedure for morphosyntactic forms (§4.4) and pragmatic functions (§4.5). Finally, the reliability of coding and different statistical procedures are described in §4.6 and §4.7.

4.2 Subjects
The data were taken from the CHILDES-database (MacWhinney, 2000). The data needed to satisfy the following criteria:

1) The transcripts had to include speech from both conversational partners, as this allows analyzing the input and tracking the different pragmatic functions related to the three pragmatic factors studied (non-specific

2) The transcripts needed to be already coded for external, non-linguistic events and context information to facilitate the coder’s analysis of the pragmatic functions. Video-recordings of the conversations would have facilitated the interpretation even more. Unfortunately, these tapes did not exist for all chilides-data selected or were not available through chilides at the time of coding. Therefore, videos could not be used in this research.

Data from three English, three Dutch and four French children were analyzed every three months between 2;0 and 3;3, that is, at six data points across this age range. For each child, a sample of 600 linguistic utterances, including minor and non-analyzable utterances (§4.3) was selected in order to achieve a comparable amount of discourse diversity across the samples of the different children. Non-linguistic utterances, such as yelling, crying or screaming were excluded from the 600-utterance-sample beforehand. For some subjects, there was no recording available at exactly the target age or there were too few utterances or references available in the recording. In that case, as many additional data as needed were used from recordings made within one month before or after the target age. In no case were there more than five weeks between the different samples of one child for a particular target age.

Since a small number of children are studied, it was important to know whether the children per language fall within the normal range of general linguistic development. For each language, the subjects’ mlu in words (mluw) was therefore calculated in the chilides-analysis tool clan (MacWhinney, 2000). The mluw of the subjects was subsequently compared to the mluw of a larger group of children of the same age and, as far as possible, of parents with the same educational level or comparable socio-economic status as the subjects (middle class or higher, see Appendix A for details of the comparison groups). The children’s mluw were converted to z-scores. The criterion for inclusion of the research subjects’ linguistic level within the normal range was set at 1 s.d. away from the mean of the comparison group. As will be shown in §4.2.1 to §4.2.3, all the children were comparable to the norm group. This makes it possible to pool the data from the different children per language in the main analyses on the children’s sensitivity to pragmatic factors in determiner and pronoun use in
Chapters 7 and 8. To make sure that the children indeed use forms in comparable ways for functions and that their data can be pooled reliably, additional statistic tests of form-function use per child will, however, also be carried out in Chapters 7 and 8. It is also important to note that in the further analyses in this study, the children were compared on the basis of age and not MLU (Chapters 5, 6, 7 and 8). It is commonly agreed that cross-linguistic comparisons based on MLU are problematic, given differences between languages in morphological complexity. Moreover, determiners, nouns, pronouns and proper names are all part of the MLU-count and a dependent variable here. A comparison based on MLU would thus be circular.

For the analysis of the input, 300 utterances of input language from the samples of each child at 2;3 and another 300 at 3;3 were selected in the same way as for the children. The two different age points made it possible to investigate possible changes in the adult input for reference as a function of the children’s increasing age and language level (§7.3.1 and §8.3.1). The language of the adult in interaction with the child was analyzed. This could be a family member (parent, grandparent) or the investigator. It is important to note that several researchers have found that the complexity of children’s language varies according to their conversational partner. For example, children use more utterances and with a longer MLU when talking to their mother than to a stranger or to older siblings (Bornstein, Haynes, Painter & Genevro, 2000; Hoff, 2008). Moreover, the nature of input may differ according to the type of interlocutor. For example, Pancsofar and Vernon-Feagans (2006) found that mothers produced longer turns and more wh-questions than fathers. It is therefore possible that the input from the different types of interlocutors studied here also varies, for example in the use of morphosyntactic forms for pragmatic functions, and that the children are influenced by this difference. Given these considerations, the possible effect of the type of interlocutor (parent or investigator) on form-function combinations in the input and child language was checked and will be reported in more detail in Chapters 7 and 8. Suffice it to say that no differences pertinent to this study were found between the parents and the investigators.

The following sections give more detailed information about the child and adult subjects per language, the data selected and the children’s language level in MLU. The subjects are a homogeneous group in terms of their family

1 The information on the children and their family situation was taken from the CHILDES-handbook (MacWhinney, 2000) and/or from the references cited for each of the corpora.
backgrounds: most of the parents have college or university degrees. The data selected are also highly comparable. All are samples from spontaneous conversations in which the children engage in daily routines and different activities, such as playing with toys (cars, trains, dolls), doing jigsaw puzzles, book reading or drinking tea. Large parts of the conversations are centered on the here-and-now. However, there are also possibilities of communicating about referents outside the here-and-now in most of the samples. For example, the conversation may be about non-present family members or friends, or children are invited to retell past experiences. In addition, the presence of an investigator in some, but not all, of the samples may create possibilities for the children to communicate about referents that are not mutually known. These conversations about the non-here-and-now and not-mutually known referents lead to discourse-new-nmk or discourse-new-mk-end references. The occurrence of these pragmatic functions enables the investigation of children’s sensitivity to the pragmatic factor of familiarity to the listener. The amount of discourse-new-nmk or discourse-new-mk-end references and the possible influence of the type of interlocutor on talk about the non-here-and-now will be investigated further in Chapter 6.

4.2.1 Dutch
Data from Abel and Matthijs (Bol, 1995) and Sarah (Van Kampen, 1994) were chosen to study the acquisition of reference in Dutch. All three children are acquiring the standard variety of Dutch spoken in the Western part of the Netherlands. This will be referred to in the rest of this thesis as ‘Dutch’.

Abel is the first-born child of university-educated parents. When Abel was 2;5, a baby-brother was born, who is sometimes present during the recordings. At 2;0, a total of 700 utterances instead of 600 were analyzed for Abel (Table 4.1). He is not very talkative in the recordings around this age and 700 utterances yielded an amount of analyzable references that was more comparable to that of the other children (more than 100, see §4.3.1). At 2;9, there were no data at all for Abel. Therefore, the first sample available, at 2;10.0, was chosen. This sample contained only 536 utterances. There were no data available within five weeks before or after 2;9 to supplement the sample. In most samples, both Abel’s mother and the investigator are present. The main interlocutor is the male investigator, whose language was coded for the input analysis at 2;3 and 3;3.
Matthijs is the first-born child of college-/university-educated parents. He has a baby-sister who is sometimes present during the recordings, but does not yet join in the conversation. At most age points, 600 utterances were available for analysis. At 2;3, an additional 25 utterances were taken from a sample recorded at 2;2.20 (Table 4.2). Both the parents and the investigator are actively engaged in the play and interaction during all of the analyzed samples. The language of the mother was coded for the input analysis, since she was the child’s main interlocutor at 2;3 and 3;3.

Sarah is the younger of two sisters. Her mother is a linguist. During the recordings, Sarah is usually alone with her mother, although her older sister sometimes interferes. The only age points for which 600 utterances were available are 2;6 and 2;9 (Table 4.3). At the other four age points, the samples...
were supplemented with utterances from transcripts as closely as possible to the target age. Sarah’s main interlocutor is her mother, whose language was therefore coded for the input analysis.

Table 4.3. Selected data per age point for Sarah and input to Sarah (Dutch)

<table>
<thead>
<tr>
<th>Age point</th>
<th>Childes-files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;0</td>
<td>sarah08</td>
<td>1;11.15</td>
<td>139</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sarah09</td>
<td>2;0.17</td>
<td>461</td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>sarah11</td>
<td>2;2.18</td>
<td>308</td>
<td>300 (mother)</td>
</tr>
<tr>
<td></td>
<td>sarah12</td>
<td>2;3.16</td>
<td>292</td>
<td></td>
</tr>
<tr>
<td>2;6</td>
<td>sarah18</td>
<td>2;6.11</td>
<td>292</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sarah19</td>
<td>2;6.18</td>
<td>308</td>
<td></td>
</tr>
<tr>
<td>2;9</td>
<td>sarah24</td>
<td>2;9.7</td>
<td>409</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sarah25</td>
<td>2;9.23</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>3;0</td>
<td>sarah28</td>
<td>2;11.27</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sarah29</td>
<td>3;0.19</td>
<td>464</td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>sarah32</td>
<td>3;3.21</td>
<td>470</td>
<td>300 (mother)</td>
</tr>
<tr>
<td></td>
<td>sarah33</td>
<td>3;4.13</td>
<td>130</td>
<td></td>
</tr>
</tbody>
</table>

The subjects’ [mlu]w was calculated for all data available in Childes at each of the target age points. Abel and Sarah show a similar [mlu]w-development, whereas Matthijs’ [mlu]w develops slightly more slowly than that of the other two children (Table 4.4). The subjects’ [mlu]w was compared to the mean [mlu]w of 13 children (longitudinal and cross-sectional data, Appendix A). Using statistical comparison (z-scores), the [mlu]w of the three Dutch subjects is for the most part comparable to the mean [mlu]w of the comparison group. Only the [mlu]w for Matthijs at 2;3 and 3;3 fall below 1 s.d. and 1.5 s.d. from the mean. It can thus be concluded that the Dutch children’s linguistic level is in general within the range of other Dutch children of the same age and with the same family background.

Table 4.4. Mean length of utterances in words ([mlu]w) per age point for the Dutch child subjects and a comparison group

<table>
<thead>
<tr>
<th>Child</th>
<th>2;0</th>
<th>2;3</th>
<th>2;6</th>
<th>2;9</th>
<th>3;0</th>
<th>3;3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abel</td>
<td>1.4</td>
<td>2.2</td>
<td>2</td>
<td>2.8</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td>Matthijs</td>
<td>1.5</td>
<td>1.6*</td>
<td>2.5</td>
<td>2.3</td>
<td>3.1</td>
<td>2.6*</td>
</tr>
<tr>
<td>Sarah</td>
<td>1.7</td>
<td>2.1</td>
<td>2.6</td>
<td>2.9</td>
<td>3.5</td>
<td>3</td>
</tr>
<tr>
<td>Mean [mlu]w comparison group</td>
<td>1.6</td>
<td>2.1</td>
<td>2.4</td>
<td>2.8</td>
<td>3.1</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Notes. * = cases in which the [mlu]w lies further than 1 s.d. from the mean of the comparison group. For details on the comparison group, see Appendix A.
4.2.2 English

Data from Adam (Brown, 1973), Nina (Suppes, 1974) and Peter (Bloom, Hood & Lightbown, 1974) were used for the study of American-English. For convenience, this will be in the rest of this thesis referred to as ‘English’.

Adam is a first-born child and his parents both have college degrees. It is worth pointing out that Adam has an Afro-American background. However, neither his speech nor the input shows highly distinctive features of African-American Vernacular English. There were no data available for Adam at 2;0. The analysis thus covers the five remaining age points between 2;3 and 3;3 (Table 4.5). Adam’s main conversational partner is his mother, whose language was coded for the input analysis at 2;3 and 3;3. In later transcripts, one of the two investigators sometimes also joins the play and conversation, although to a lesser extent than the mother.

<table>
<thead>
<tr>
<th>Age point</th>
<th>CHILDES-files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;0</td>
<td>no data available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>adamo1</td>
<td>2;3.4</td>
<td>600</td>
<td>294 (mother)</td>
</tr>
<tr>
<td>2;6</td>
<td>adamo7</td>
<td>2;6.3</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>2;9</td>
<td>adami3</td>
<td>2;9.4</td>
<td>546</td>
<td></td>
</tr>
<tr>
<td></td>
<td>adami4</td>
<td>2;9.18</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>3;0</td>
<td>adamo20</td>
<td>3;0.10</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>adamo26</td>
<td>3;3.4</td>
<td>600</td>
<td>300 (mother)</td>
</tr>
</tbody>
</table>

Nina is a first-born child and lives alone with her mother, who is a linguist. There were no data for Nina at 2;6 exactly. Therefore, the closest data point available has been analyzed, which was 2;5.27 (Table 4.6). A difference of only three days from the desired age point should not influence the data significantly. Nina’s mother is the only interlocutor in the samples selected. Her language was thus coded for the input analysis at 2;3 and 3;3.

2 The CHILDES-handbook does not offer information about Nina’s family situation. The description here is based on the extra-linguistic information in the transcripts.
Table 4.6. Selected data per age point for Nina and input to Nina (English)

<table>
<thead>
<tr>
<th>Age point</th>
<th>childes-files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;0</td>
<td>nina07</td>
<td>2;0.24</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>nina16</td>
<td>2;3.5</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>2;6</td>
<td>nina30</td>
<td>2;5.27</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>2;9</td>
<td>nina32</td>
<td>2;9.13</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>3;0</td>
<td>nina42</td>
<td>3;0.3</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>nina55</td>
<td>3;3.8</td>
<td>600</td>
<td>300 (mother)</td>
</tr>
</tbody>
</table>

Peter is the first-born child of college-educated parents. He has a baby sister, who is present during some of the selected recordings, but does not yet talk. There were no data available for Peter at 3;0. In the last available transcript, he is 3;1.20. This transcript was included in the analysis for age point 3;3 (Table 4.7). Peter’s parents are present during all the selected recordings. However, the child primarily plays and interacts with the two investigators, whose language was therefore investigated in the input analysis. At 2;3, investigator Patsy was the main interlocutor and a sample of 300 utterances of her language was analyzed. At 3;3, Peter first played with Patsy and later with the other investigator, Lois. Therefore, 250 input utterances from Patsy were analyzed and a remaining 50 from Lois.

Table 4.7. Selected data per age point for Peter and input to Peter (English)

<table>
<thead>
<tr>
<th>Age point</th>
<th>childes-files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;0</td>
<td>peter06</td>
<td>2;0.10</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>peter10</td>
<td>2;3.3</td>
<td>600</td>
<td>300 (investigator)</td>
</tr>
<tr>
<td>2;6</td>
<td>peter15</td>
<td>2;6.16</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>2;9</td>
<td>peter18</td>
<td>2;9.15</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>no data available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>peter20</td>
<td>3;1.20</td>
<td>600</td>
<td>300 (investigator)</td>
</tr>
</tbody>
</table>

The \textit{mlu}w indicates that language development is slower for Adam than for Peter and Nina (Table 4.8). The subjects’ \textit{mlu}w was compared to the mean \textit{mlu}w of 22 children (Appendix A). Only three scores of the sixteen data points from all three subjects lie between 1 s.d. and 1.5 s.d. from the mean: Peter at 2;6 and 2;9 and Nina at 2;3 (Table 4.8). The language level of all three subjects is, therefore, overall within the range of what could be expected at their age for children with a comparable family background. The three English children can therefore function as representatives of this child population and can also be compared to each other.
Table 4.8. Mean length of utterances in words (\textit{mlu}w) per age point for the English child subjects and a comparison group

<table>
<thead>
<tr>
<th>Child</th>
<th>2;0</th>
<th>2;3</th>
<th>2;6</th>
<th>2;9</th>
<th>3;0</th>
<th>3;3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>no data</td>
<td>2.1</td>
<td>2.7</td>
<td>2.4</td>
<td>3.6</td>
<td>3.8</td>
</tr>
<tr>
<td>Nina</td>
<td>2.1</td>
<td>3.2(^a)</td>
<td>3</td>
<td>3</td>
<td>3.5</td>
<td>3.7</td>
</tr>
<tr>
<td>Peter</td>
<td>2.3</td>
<td>2.5</td>
<td>3.6(^a)</td>
<td>3.5(^a)</td>
<td>no data</td>
<td>3.4</td>
</tr>
<tr>
<td>Mean \textit{mluw} comparison group</td>
<td>2</td>
<td>2.4</td>
<td>2.7</td>
<td>2.8</td>
<td>3.6</td>
<td>3.9</td>
</tr>
</tbody>
</table>

Notes. \(^a\) = cases in which the \textit{mluw} lies further than 1 s.d. from the mean of the comparison group. For details on the comparison group, see Appendix A.

4.2.3 French

At the time of coding, there were longitudinal data available for only two children acquiring French in the desired age range: Anne (De Cat & Plunkett, 2002) and Philippe (Suppes, Smith & Leveillé, 1973). Since there was no third French child available in childes whose data covered the total age range 2;0-3;3, a mixed longitudinal-cross-sectional design was used to obtain enough data: Grégoire’s data range from 2;0-2;6 and Léa’s data range from 2;9-3;3 (De Cat & Plunkett, 2002). Léa is acquiring the Belgian variety of French.

There are some dialectal differences between French spoken in France and in Belgium, for example in the use of imperatives, complementizers and modal verbs. There are no differences reported on the use of morphosyntactic forms for pragmatic functions. For that reason, the use of a corpus from Belgium was not seen as a problem for this study.\(^3\) Both varieties will be referred to as ‘French’ in this thesis.

Anne is a first-born child whose baby-brother was born when she was 2;6.

Her mother is a landscape architect and her father was in temporary work in show business at the time of recording. At all ages except 2;6 and 3;0, the sample had to be supplemented with utterances from transcripts close to the target age to collect the desired 600 utterances (Table 4.9). At 2;0, only 515 utterances could be collected. In the samples at 2;3 and 3;3, Anne’s main interlocutor is the investigator, whose language was therefore coded for the input analysis.

\(^3\) The York-corpus contains longitudinal data for a third child between 1;9 and 3;2: Max. This child, however, is a speaker of Canadian-French. There are several syntactic and lexical differences between Canadian-French and Standard French. Some of these affect the use of determiners and pronouns, which are important for the current study. As the coder was a second-language speaker of Standard French, it was thought to be wise not to include a corpus with many dialectal traits.
<table>
<thead>
<tr>
<th>Age point</th>
<th>chil-des-files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;0</td>
<td>para004</td>
<td>1;11.29</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td></td>
<td>para005</td>
<td>2;0.13</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>para006</td>
<td>2;0.27</td>
<td>224</td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>para010</td>
<td>2;2.30</td>
<td>223</td>
<td>300 (investigator)</td>
</tr>
<tr>
<td></td>
<td>para011</td>
<td>2;3.25</td>
<td>144</td>
<td></td>
</tr>
<tr>
<td></td>
<td>para012</td>
<td>2;4.2</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>2;6</td>
<td>para016</td>
<td>2;6.2</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td></td>
<td>para017</td>
<td>2;6.18</td>
<td>285</td>
<td></td>
</tr>
<tr>
<td>2;9</td>
<td>para021</td>
<td>2;8.20</td>
<td>345</td>
<td></td>
</tr>
<tr>
<td></td>
<td>para022</td>
<td>2;9.15</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>3;0</td>
<td>para025</td>
<td>3;0.2</td>
<td>419</td>
<td></td>
</tr>
<tr>
<td></td>
<td>para026</td>
<td>3;0.10</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>para031</td>
<td>3;2.29</td>
<td>278</td>
<td>300 (investigator)</td>
</tr>
<tr>
<td></td>
<td>para032</td>
<td>3;3.16</td>
<td>322</td>
<td></td>
</tr>
</tbody>
</table>

Philippe is the first-born child of university-educated parents. Since there were no data available at 2;0, the investigation covers the period from 2;3-3;3 (see Table 4.10). The ages at which there were not enough utterances available are 2;9 and 3;3. For these age points, the data were supplemented with utterances from recordings that were closest to the target age. Both the parents and the investigator are present during the recordings of the analyzed samples. At 2;3, Philippe’s mother is the main interlocutor, but there were only 290 utterances from her available in the sample. Therefore, ten utterances from the father had to be added to complete the sample of 300 utterances for the input analysis. At 3;3, the child interacts mostly with the investigator, whose language was coded for the input analysis at this age point.
Table 4.10. Selected data per age point for Philippe and input to Philippe (French)

<table>
<thead>
<tr>
<th>Age point</th>
<th>Child files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;0</td>
<td>no data available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>phil07</td>
<td>2;3.0</td>
<td>383</td>
<td>290 (mother) and 10 (father)</td>
</tr>
<tr>
<td></td>
<td>phil08</td>
<td>2;3.7</td>
<td>217</td>
<td></td>
</tr>
<tr>
<td>2;6</td>
<td>phil11</td>
<td>2;6.13</td>
<td>489</td>
<td></td>
</tr>
<tr>
<td></td>
<td>phil12</td>
<td>2;6.20</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>2;9</td>
<td>phil21</td>
<td>2;8.29</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td></td>
<td>phil22</td>
<td>2;9.15</td>
<td>332</td>
<td></td>
</tr>
<tr>
<td>3;0</td>
<td>phil28</td>
<td>3;0.6</td>
<td>563</td>
<td></td>
</tr>
<tr>
<td></td>
<td>phil29</td>
<td>3;0.20</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>phil32</td>
<td>3;2.29</td>
<td>116</td>
<td>300 (investigator)</td>
</tr>
<tr>
<td></td>
<td>phil33</td>
<td>3;3.12</td>
<td>484</td>
<td></td>
</tr>
</tbody>
</table>

As mentioned above, data from both Grégoire and Léa were analyzed to cover the age range under investigation. Data are available from Grégoire at 2;0, 2;3 and 2;6 and from Léa at 2;9, 3;0 and 3;3.

Grégoire is the third-born child of college-educated parents. The sample at 2;0 had to be supplemented with data from 1;11 in order to collect 600 utterances (Table 4.11). For the sample at 2;3, only 438 utterances were available. The closest age point is 2;1.24. This is more than a month from the target age and for that reason, it was chosen not to supplement the sample at 2;3 with utterances from 2;1. At 2;6, there were no data available at all. Consequently, the closest available data from 2;5 were analyzed. Grégoire’s mother and the investigator are always present during the recordings. Grégoire’s two older brothers are sometimes present and participate in the interaction on those occasions. At 2;3, Grégoire’s mother is the main conversational partner. There were, however, only 270 utterances of the mother available in the sample. The input data were therefore supplemented with 30 utterances from the investigator.
Table 4.11. Selected data per age point for Grégoire and input to Grégoire (French)

<table>
<thead>
<tr>
<th>Age point</th>
<th>CHILDES-files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;0</td>
<td>greg04</td>
<td>1;11.21</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td></td>
<td>greg05</td>
<td>2;0.5</td>
<td>466</td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>greg07</td>
<td>2;3.1</td>
<td>438</td>
<td>270 (mother) and 30 (investigator)</td>
</tr>
<tr>
<td>2;6</td>
<td>greg09</td>
<td>2;5.13</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>greg10</td>
<td>2;5.27</td>
<td>476</td>
<td></td>
</tr>
</tbody>
</table>

Léa is the first-born child of college-educated parents. At all target age points, there were 600 utterances from Léa available (Table 4.12). No investigator was involved in the recordings, which were conducted by Léa’s grandmother. The child’s mother and grandfather are also sometimes present. The grandmother is the child’s main adult interlocutor at 3;3. Her language production was therefore included in the input analysis.

Table 4.12. Selected data per age point for Léa and input to Léa (French)

<table>
<thead>
<tr>
<th>Age point</th>
<th>CHILDES-files</th>
<th>Exact age</th>
<th>Number of analyzed utterances (child)</th>
<th>Number of analyzed utterances (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;9</td>
<td>liea002</td>
<td>2;9.5</td>
<td>357</td>
<td></td>
</tr>
<tr>
<td></td>
<td>liea003</td>
<td>2;9.21</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>3;0</td>
<td>liea008</td>
<td>3;0.5</td>
<td>362</td>
<td></td>
</tr>
<tr>
<td></td>
<td>liea009</td>
<td>3;0.24</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td>liea012</td>
<td>3;3.8</td>
<td>286</td>
<td>300 (grandmother)</td>
</tr>
<tr>
<td></td>
<td>liea013</td>
<td>3;3.25</td>
<td>314</td>
<td></td>
</tr>
</tbody>
</table>

The linguistic level of the four French children can be compared on the basis of the \( \text{mlu}_w \). Philippe turns out to be most advanced, followed by Grégoire and Anne, who have comparable \( \text{mlu}_w \)’s at 2;0 and 2;3 (Table 4.13). Léa’s \( \text{mlu}_w \) at 2;9 fits with Grégoire’s \( \text{mlu}_w \) at 2;6. From 2;9 onwards, Léa has a slightly higher \( \text{mlu}_w \) than Anne and is thus seen as linguistically a little ahead of Anne. The French comparison group consists of ten longitudinally recorded children from parents with comparable educational backgrounds or socio-economic status as the subjects (Appendix A). The French subjects’ \( \text{mlu}_w \) can be evaluated against that of the comparison group up to 3;0. At 3;3, too few data from the comparison group (n=1) were available to make any result valid. Only four scores of the seventeen data points from all four subjects lie between 1 s.d. and 1.5 s.d. from the mean: Grégoire at 2;6, Philippe at 2;3 and 3;0 and Léa at 3;0. The French children’s linguistic level is therefore
largely comparable to that of other children of the same age and from the same family background.\textsuperscript{4}

Table 4.13. Mean length of utterances in words (\textit{mlu}w) per age point for the French child subjects and a comparison group

<table>
<thead>
<tr>
<th>Child</th>
<th>2;0</th>
<th>2;3</th>
<th>2;6</th>
<th>2;9</th>
<th>3;0</th>
<th>3;3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>2.5</td>
<td>2.4</td>
<td>3.3</td>
<td>3.6</td>
<td>3.8</td>
<td>4.3</td>
</tr>
<tr>
<td>Grégoire</td>
<td>2.1</td>
<td>2.6</td>
<td>4.3*</td>
<td>no data</td>
<td>no data</td>
<td>no data</td>
</tr>
<tr>
<td>Léa</td>
<td>no data</td>
<td>no data</td>
<td>no data</td>
<td>4</td>
<td>4.1*</td>
<td>4.3</td>
</tr>
<tr>
<td>Philippe</td>
<td>no data</td>
<td>3.4*</td>
<td>3.8</td>
<td>4</td>
<td>4.2*</td>
<td>4.7</td>
</tr>
<tr>
<td>Mean \textit{mluw} comparison group</td>
<td>2.1</td>
<td>2.6</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Notes. \textsuperscript{*} = cases in which the \textit{mluw} lies further than 1 s.d. from the mean of the comparison group. For details on the comparison group, see Appendix A.

4.3 Data selection

Not all of the child and adult utterances in the selected samples could be analyzed for person or object reference. Some did not contain a person or object reference at all, whereas others did contain one or more referential expressions to persons or objects, but these utterances (and referential expressions) could not be analyzed for a variety of reasons. The non-analyzable utterances are discussed in §4.3.1. The demarcation of person and object reference, the focus of the current study, is discussed in §4.3.2. The same analysis procedure was used for child and input language, but the examples discussed here will come primarily from the child data.

4.3.1 Non-analyzable utterances

From the 600 utterances per child sample and 300 per input sample, several utterances were excluded from the analysis. First, some utterances did not contain a person or object reference at all, as in (1). Others consisted only of minimal responses, such as \textit{hmm}, elliptic answers to yes/no questions (see 2), onomatopoeia or songs. Utterances that were partly or wholly unintelligible, as in (3), were also excluded, even if the utterance contained a clear person or object reference (see 4). The pragmatic function of the referent could not be interpreted.

\textsuperscript{4} It is interesting to note that family background and socio-economic status are clearly important, since a comparison of the French subjects’ \textit{mluw} with the mean \textit{mluw} of a sample of 40 subjects (longitudinal and cross-sectional) collected by Marie-Thérèse Le Normand did reveal differences. The z-scores of the four French subjects fall 1 s.d. above the mean of le Normand’s sample in 65\% of the cases. Le Normand’s sample can, however, not be directly compared with the subjects in this research, since half of her subjects are from lower socio-economic backgrounds. As Hoff and Tian (2005) have indicated, low socio-economic background and lower maternal education are related to slower language development.
reliably in such cases. Unfinished utterances, as in (5), were left out for a similar reason: the pragmatic function of the referent was most often unclear, since the coder could not know how the speaker intended to finish her utterance to possibly clarify the reference.

(1) **No person or object reference (Philippe, 2;6, French)**

FAT: Et le bateau?

‘And the boat?’

FAT: De quelle couleur il est le bateau?

‘What color is the boat?’

CHI: Rouge.

‘Red’

> excluded

(2) **Elliptic yes/no answer (Peter, 2;6, English)**

INV: Did you come on the train?

CHI: Yeah.

> excluded

(3) **Partly and wholly unintelligible utterances without reference (Abel, 2;0, Dutch)**

MOT: Nou is tie kapot.

Now it’s broke’

CHI: xx! > excluded

%com: Unintelligible

CHI: xx weg.

‘xx gone’

> excluded

(4) **Partly unintelligible utterance with intelligible reference (Peter, 2;6, English)**

CHI: What’s this?

INV: I think that’s a daddy cow # Peter.

CHI: xxx daddy cow. > excluded

(5) **Referent in unfinished utterance (Anne, 2;6, French)**

CHI: Et ça, je range +/

‘And that, I order...’

CHI: Pourquoi ç(e) est +//?

‘Why is it...’

> excluded

Utterances with a referential expression that were (almost) verbatim imitations of an immediately preceding utterance, as in (6), were also excluded. These repeated utterances possibly contain morphosyntactic devices and form-function combinations that are more complex or advanced than the ones that
the children can produce spontaneously. In the adult data, these imitations merely function as clarifications and were also excluded.

(6) **Imitation (Peter, 2;0, English)**

**INV:** And smoke over there.
**CHI:** Smoke o(ver) there. > excluded

Self-repetitions of utterances with a referential expression, as in (7), were also excluded. Self-repetitions were defined as utterances in which the speaker repeats a referential form verbatim one or more times with the same form-function combination and with no more than three intervening utterances from the conversational partner. In these cases, only the first utterance was analyzed. Including all self-repetitions might have led to an overrepresentation of certain form-function combinations in the analysis. For that reason, frequent repetitions of the same morphosyntactic form for different referents with intervening child utterances were also coded as self-repetitions after ten times and subsequently excluded. This last type of self-repetition occurred sporadically. The most prominent example occurred in the data of the Dutch child Sarah at 3;3. She wanted her mother to name the numbers on the clock by asking her *en deze?* ‘and this one?’ over and over again.

(7) **Self repetition (Grégoire, 2;0, French)**

**CHI:** La poussette de Grégoire.
‘Grégoire’s stroller’

**INV:** Oui.
‘Yes’

**INV:** Ah!
‘Oh!’

**CHI:** La poussette de Grégoire.
‘Grégoire’s stroller’ > excluded

**INV:** Oui oui oui oui la poussette de Grégoire.
‘Yes yes yes yes Grégoire’s stroller’

The total percentage of non-analyzable utterances, both with and without a person or object reference, varies across the subjects and across different

---

5 Excluding imitations from the current analysis does not at all mean that these utterances are considered useless in the acquisition process. According to Ninio and Snow (1996: 56-57), an imitation is meaningful communicative behavior, which is part of the larger picture of language acquisition.
samples. It generally ranges between 45%-60% of all utterances in the child data and between 30%-50% in the input data. This may seem high, but it is apparently quite usual in spontaneous conversations with children of this age: the children studied here form a homogeneous group and they engage in normal activities and daily routines. Most of the non-analyzable utterances are minors or other utterances that do not contain a referential expression at all.

4.3.2 Selection of analyzable references

In the selected utterances, not all nouns, pronouns or proper names were analyzed. The current study focuses on third person reference with nouns, pronouns and proper names to persons and concrete objects. References to mass substances, such as ‘coffee’, ‘sugar’ or ‘candy’, and body parts, such as ‘arm’ or ‘head’, were also included. A pilot analysis revealed that these were frequent in the transcripts. Henceforth, these are classed under ‘objects’. All person or object references in the same utterance were analyzed separately. Similarly, the different referential expressions in complex constructions, such as one of these, were also analyzed separately.

Person or object references that were part of a false start were not analyzed. References to generic locations such as ‘kitchen’, ‘floor’, ‘home’, or ‘playground’ in example (8), were excluded.

(8) Reference to generic location (Matthijs, 3;3, Dutch)

MOT: Volgens mij heb je nog helemaal niet aan Evelien verteld waar jij geweest bent met Kiko.
*I think you haven’t told Evelien yet where you went with Kiko (=kindergarten).*

CHI: Ja.
*Yeah*

MOT: Waar ben # +/.
*Where have...*

CHI: Speeltuin. > excluded
*Playground*

In distinguishing reference to persons and objects from reference to locations, this research follows Hickmann (2003). Spatial reference is seen as a special type of reference, involving other linguistic devices than the ones investigated in this study, for example spatial prepositions. The excluded references to generic locations were not very frequent, on average
5% of the nouns, pronouns and proper names used in the child data and 7% in the input data. If, however, a person or object reference was used in a phrase that served the semantic function of 'location', but was not a location in itself, for example the table in (9), the reference was not excluded from the analysis.

(9) Analyzable object reference (Nina, 3:0, English)

\text{MOT: Where you putting the cake?}
\text{CHI: Right there.}
\text{MOT: On what?}
\text{CHI: On the table.}

Bare nouns referring to persons or objects that are part of common verbal constructions, such as go by car or aller en voiture, were also excluded. Nouns phrases in such constructions show idiosyncratic behavior with regard to the presence of the determiner. These constructions were not very frequent in the data (less than 0.5% of all person and object references in the child and adult data) Moreover, nouns that occurred in negations (10), quantificational constructions (11) and nouns that were preceded by a wh-word (12) were not analyzed. First of all, it is hard to establish to which referents these forms actually refer (Avrutin, 1999: 26). Second, the frequency of negated, quantified and questioned nouns was very low in the data, on average less than 3% of all person and object references in the child data and 4% in the adult data.

(10) Negated noun (Nina, 3:3, English)

\text{MOT: Which clothes do you want to go on her now?}
\text{CHI: No clothes.} > excluded

(11) Quantified noun (Philippe, 2:6, French)

\text{CHI: Beaucoup de crayons là-dedans.} > excluded
‘A lot of pencils in there’

(12) Wh-word and noun (Sarah, 3:3, Dutch)

\text{CHI: Welk boekje hebbe(n) we allemaal?} > excluded
‘Which book do we have?’

\text{6 The three languages differ in the realization of these constructions. Thus, the Dutch translation of go by car, which is met de auto gaan, does not involve a bare noun. In the case of Dutch, the determiner-noun combination was included in this study.}
The child utterance *screw* in (13) is morphosyntactically ambiguous between a nominal or verbal reading. Potential referential expressions of this type were also excluded, since their morphosyntactic, and therefore pragmatic, interpretation is unclear. Some nominal forms were excluded because of semantic ambiguity, as in (14). The child uses a word without a meaning and therefore without a referent in the adult language.

(13) Ambiguous morphosyntax (Peter, 2;0, English)

inv: What is that [= screwdriver]?  
chi: Screw? > excluded

(14) Ambiguous semantics (Léa, 2;9, French)

chi: Et elle est où, la tata? > excluded
‘And where’s the tata?’
%com: Just asking (something meaningless, i.e. ‘the tata’) for the sake of asking something.

grm: La tata?  
‘The tata’
grm: Ça est qui, ça, la tata?  
‘What’s that, ‘the tata’?’

chi: Non, la cada. > excluded
‘No, the cada’
%com: Word probably invented for the occasion: she is just being difficult.

In sum, from the 600 child utterances selected per age point, around 300 could be analyzed for person/object reference (§4.3.1). These utterances yielded 100 to 300 referential expressions to third person entities and objects at the early age points. The number of referential expressions increases up to 200-480 at 3;3 (Table 4.14). In the input, 150 out of the 300 selected utterances per age point could be analyzed for person and object reference. This yielded 100 to 350 referential expressions per age point. In the following sections, the analysis of the morphosyntactic forms (§4.4) and pragmatic functions (§4.5) of the analyzable person and object references are discussed.
### Table 4.14. Overview of raw number of analyzable person and object references (tokens) per child per age point in the child and input data

<table>
<thead>
<tr>
<th>Language</th>
<th>Child</th>
<th>2;0</th>
<th>2;3</th>
<th>2;6</th>
<th>2;9</th>
<th>3;0</th>
<th>3;3</th>
<th>Input 2;3</th>
<th>Input 3;3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>Abel</td>
<td>123</td>
<td>140</td>
<td>188</td>
<td>215</td>
<td>153</td>
<td>225</td>
<td>121</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Matthijs</td>
<td>170</td>
<td>229</td>
<td>271</td>
<td>260</td>
<td>294</td>
<td>191</td>
<td>179</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>Sarah</td>
<td>218</td>
<td>264</td>
<td>237</td>
<td>334</td>
<td>311</td>
<td>259</td>
<td>182</td>
<td>235</td>
</tr>
<tr>
<td>English</td>
<td>Adam</td>
<td>nd</td>
<td>289</td>
<td>286</td>
<td>313</td>
<td>357</td>
<td>290</td>
<td>204</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>Nina</td>
<td>299</td>
<td>397</td>
<td>408</td>
<td>280</td>
<td>446</td>
<td>486</td>
<td>341</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>Peter</td>
<td>159</td>
<td>247</td>
<td>321</td>
<td>296</td>
<td>nd</td>
<td>461</td>
<td>179</td>
<td>200</td>
</tr>
<tr>
<td>French</td>
<td>Anne</td>
<td>113</td>
<td>193</td>
<td>329</td>
<td>397</td>
<td>283</td>
<td>418</td>
<td>213</td>
<td>212</td>
</tr>
<tr>
<td></td>
<td>Grégoire</td>
<td>180</td>
<td>155</td>
<td>321</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>217</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td>Léa</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>249</td>
<td>270</td>
<td>235</td>
<td>nd</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>Philippe</td>
<td>nd</td>
<td>302</td>
<td>294</td>
<td>299</td>
<td>457</td>
<td>402</td>
<td>212</td>
<td>160</td>
</tr>
</tbody>
</table>

Notes. nd = no data available.

### 4.4 Coding of morphosyntactic forms for reference

This research focuses on reference with nouns and different types of determiners and on reference with pronouns as opposed to full lexical nouns and proper names. Only overt forms were coded and included in the analysis. Non-realized arguments, that is, where the speaker leaves out one or more of the obligatory arguments of a predicate, were not coded nor analyzed (§1.1). By coding and analyzing only overt forms, no knowledge of argument structure is attributed to the children where this is not evident in their language production (Laakso & Smith, 2007). In §4.4.1, an overview is given of the different types of nouns and determiners coded. Different pronominal forms are discussed in §4.4.2. Information on the coding of proper names is presented in §4.4.3. Finally, §4.4.4 discusses the coding of referential expressions that occur in dislocations or it-clefts (see also §2.2.3).

#### 4.4.1 Determiners and nouns

In the analysis of nominal forms for reference (Chapter 7), the main focus will be on the noun and type of determiner used. The following determiner types were coded and occur in all three languages (see also §2.2.1.4): singular indefinite determiners (e.g. *een konijn/a rabbit/un lapin*), definite determiners (e.g. *het konijn/the rabbit/le lapin*), demonstrative determiners (e.g. *dit konijn/this rabbit/ce lapin*), possessive determiners (e.g. *mijn konijn/my rabbit/mon lapin*) and numeral determiners (e.g. *twee konijnen/two rabbits/deux lapins*). The form of the Dutch and French singular indefinite determiner (*een* and *un/une*) and the numeral ‘one’ in these languages are orthographically identical. On the basis of
the context, the decision was made as to which category the form belonged. Two-word utterances that consisted only of a demonstrative and noun, as in (15), were analyzed carefully. As Van Kampen (2005) pointed out, such a form might very well be a sequence of a demonstrative pronoun followed by a noun, where the pronoun functions as a pragmatic operator to attract attention. On the basis of the context and the interlocutor’s interpretation, it was decided on either analysis as a pronoun and ungrammatical bare noun or analysis as a combination of a demonstrative determiner and noun. For instance in (15), the child gives the investigator a cookie while uttering \textit{die mondje}. It is plausible that the shared focus of attention is on the cookie. The investigator indeed interprets the child’s use of the demonstrative \textit{die} as referring to ‘cookie’ and not to ‘mouth’. This interpretation was subsequently followed in coding.

\begin{tabular}{llll}
\hline
\textbf{15} & Sequence of demonstrative pronoun and noun (Matthijs, 2;0, Dutch) \\
\hline
\textit{CHI}: & \textit{Die mondje}. \\
& ‘That (must go in your) mouth’ \\
\textit{%act}: & Gives piece of cookie to the investigator. \\
\textit{INV}: & \textit{Moet ik die in m(i)jn mondje doen?} \\
& ‘Do I have to put that one in my mouth?’ \\
\hline
\end{tabular}

As discussed in §2.2.1.4, some nominal constructions for reference do not exist in all three languages studied. The genitive construction does not exist in French and this construction was therefore only coded in the Dutch and English data (e.g. \textit{Sophie’s konijn}/\textit{Sophie’s rabbit}). It is important to remember that there are always two references in a genitive construction: the possessor and the possessed item. Both were coded morphosyntactically and pragmatically (see §4.5 for pragmatic coding). For instance, the referential expression \textit{head} in (16) was coded as the genitive noun and \textit{this dolly’s} was coded separately as a noun with a demonstrative determiner. At the younger ages, the children sometimes produced sequences of bare nouns or sequences consisting of a proper name and a bare noun, as in (17). This construction could very well serve the semantic function of possession in some utterances. Since the genitive -s is missing, however, the form \textit{checker} in (17) was not coded as a genitive but as an (ungrammatical) bare noun (see below in this section).

7 In Dutch, there is a phonetic difference between the use of \textit{een} as an indefinite determiner and as a numeral. That is, as an indefinite determiner, it is pronounced with a schwa (§2.2.1.1). If this difference was indicated in the data, the transcriber’s decision was followed in coding.
(16) Genitive construction (Nina, 3;3, English)

chi: I’m gonna put this on this dolly’s head.

(17) Bare noun with possible interpretation of possession (Adam, 2;3, English)

inv: Are those your checkers?

chi: Adam checker.

There are differences between Dutch, English and French with regard to the obligatory use of determiners (§2.2.1.4). This has an impact on the coding procedure. Mass nouns (e.g. suiker/sugar) and indefinite plural nouns (e.g. konijnen/rabbits) are expressed by means of a bare noun in Dutch and English. These forms were coded separately as ‘grammatical bare noun’ and ‘indefinite plural as grammatical bare noun’ in these languages. In French, indefinite plural nouns are preceded by the plural determiner (e.g. des lapins) and mass nouns by the partitive determiner (e.g. du sucre). These determiner categories were therefore added in the coding of the French data.

There were two types of nominal forms that occurred mostly in child language and hardly in the input. Firstly, in the early stages of language acquisition, the children often produce nouns without a determiner where it is obligatory in the adult language, as in (18) (see also §3.2.1). These forms were coded as ‘ungrammatical bare nouns’. As pointed out above, bare nouns are grammatical for indefinite plural and mass nouns in Dutch and English (§2.2.1). These forms were coded in the advantage of the child. Thus, if an indefinite plural or mass interpretation of the referent was possible, it was coded as such. Only if the referent was already specific and introduced in discourse and the child still used a bare form, was it coded as an ungrammatical bare noun. Moreover, in sporadic cases, the adult prompted the child to produce a label by giving the determiner, as in (19). The child only needed to complete the utterance with a bare noun. In this case, the bare noun was not coded as ungrammatical, but as a grammatical bare noun.

(18) Ungrammatical bare noun (Adam, 2;3, English)

chi: Hit ball.

(19) Grammatical bare noun in completing the interlocutor’s utterance (Anne, 2;9, French)

%sit: Book reading.

inv: Il pleut donc il prend son +/.

‘It’s raining, so he takes his…’
Secondly, at the early age points, the children produced nouns preceded by a filler syllable, as in (20). This form results from a very common process in determiner acquisition: young children produce a filler syllable before the noun instead of a full determiner (§3.2.1).

(20) Noun preceded by filler syllable (Anne, 2;6, French)

CHI: Ohlala e bateau là.

‘Oh a boat there’

Orthographic transcriptions before nouns of /ee[n] in the Dutch data and e/a in the French data were coded as fillers. Grégoire’s data contain an additional phonological transcription\(^8\) that provides information on the production of determiners or fillers. This phonological information was used in coding, since the orthographic transcription does not discriminate between fillers and full determiners in the data for this child. It is important to note that in English, filler syllables have the same phonological form as the indefinite determiner. They can therefore not be distinguished from each other in this language. This introduces a bias for the English data. The English children might use the indefinite determiner differently from the French and Dutch children at least in the early stages, that is, when the definite determiner has not yet been fully acquired. This point will be addressed in the analysis of determiner use for pragmatic functions in Chapter 7.

Three additional remarks are necessary to further explain the coding of determiners and nouns. Firstly, nouns and determiners were not coded nor analyzed for number and gender, with the exception of the separate categories for indefinite singular nouns and plural nouns. These forms differ from each other with regard to determiner realization in Dutch and English (see above). Secondly, as the examples in (21) show, the subjects sometimes produced noun phrases where the determiner was followed by an adjective and the position of the noun was either empty (in Dutch and French) or filled with the dummy element one (in English). In these elliptic

---

\(^8\) Actually, there appear to circulate two versions of the Grégoire-data among researchers. The version on childes at the time of coding only contained an orthographic transcription. An older version contained an additional phonological transcription. I would like to thank Maren Pannemann for bringing this to my attention and Elisabeth van der Linden for providing me the data with the phonological transcription.
constructions, the noun is omitted, since the speaker assumes that the listener is familiar enough with the referent to identify it without being explicitly mentioned. Morphosyntactically, these elliptic noun phrases were coded and analyzed according to the type of determiner used. Adjectives that were not preceded by a determiner nor followed by a noun, for example rode ‘red’, were excluded. These forms can either be the result of determiner and noun omission or be intended as adjectives only. Their status within the analysis of nominal reference is therefore unclear.

(21) Elliptic noun phrases in which the category noun is empty.

a. mot: Welke heks?
   ‘Which witch?’
   chI: Die rode.
       (Sarah, 2;9, Dutch)
       ‘That red (one)’

b. chI: Right there’s the yellow one # right there.
       (Peter, 2;9, English)

c. chI: Je veux voir l’autre.
       (Philippe, 2;9, French)
       ‘I would like to see the other (one)’

Thirdly, the CHILDES transcription system CHAT contains a special marker [?] to indicate that the transcriber is not completely sure about the realized form and that the best guess has been transcribed (MacWhinney, 2000). Forms that were labeled as such by the transcribers have been coded as regular forms. The reason for this is that the transcribers of some data appeared to make more use of this symbol than others. In the CHILDES-data, one can unfortunately not be sure that the transcribers of different corpora applied the same criteria.

The analysis of nouns and determiners for reference focuses on the most frequent determiner types, that is, singular indefinite, definite, demonstrative and possessive determiners (see also §7.2). These determiner types also show the sharpest contrast in terms of the cognitive status they convey and the appropriate or inappropriate use for pragmatic functions (§2.4.1). The choice for the singular indefinite determiner was made for comparability, since only French has plural indefinite determiners. The other determiner types, which are relatively infrequent (§5.3.1), will also be included in the analysis, but in a combined category ‘other’, together with forms that do not occur in all three languages. This decision was made to avoid categories with few items, which is disadvantageous for the statistical analysis (§4.7). The category ‘other’ will be discussed in more detail in §7.2. The elaborate coding of different types of determiners and nominal forms as described here made it possible to investigate the behavior of the category ‘other’ in relation to pragmatic functions in more detail where necessary.
4.4.2 Pronouns

The use of pronominal forms for pragmatic functions will be compared to the use of full lexical nouns, that is, all forms described in §4.4.1 together, and proper names (§4.4.3). The coding of pronouns comprises the third person pronominal forms described in §2.2.2. A distinction was made between personal, demonstrative, possessive, relative and reflexive pronouns. Numerals used without a noun (see 22) were also included and will be classed under the category pronouns.

(22) Numeral (Léa, 2;9, French)

%act: The child offers her grandmother a sweet.

chi: T(u) en veux une?
'Do you want one?'

Personal pronouns can occur in their subject form, as in (23), or in their object form, as in (24) (§2.2.2). Both subject forms and object forms were coded, but no distinction between the two was made in the further analyses. The same applies to the distinction between strong and weak/clitic pronouns. For example, the strong demonstrative pronoun ça in (25) was coded and analyzed in a similar way as the weak form ce in the same example.

(23) Subject form of personal pronoun (Adam, 2;9, English)

chi: She doing # on table.

(24) Object form of personal pronoun (Nina, 2;0, English)

mot: That’s a monkey puppet.

chi: Hold him.

(25) Strong and clitic form of demonstrative pronoun (Léa and input to Léa, 3;0, French)

grm: Qu’est-ce que c(e) est, ça?
‘What’s that?’

chi: Ce sont des ciseaux.
‘Those are scissors’

Possessive pronouns can be used predicatively, as in (26), or attributively, as in (27), where they function as a determiner. As has been explained in §2.2.2, the possessive determiner clearly refers to a person or object itself. It was therefore included separately in the analysis of pronouns. Thus, in (27), the full construction sa chèvre is coded as a noun with a possessive determiner and the possessive sa is also analyzed separately as a possessive pronoun.
Third person pronominal forms can carry number and gender in the languages studied, but this was not coded. Their role in the acquisition of reference does not fall within the scope of this thesis (§1.1 and §2.2.2).

In the analyses of pronoun use for pragmatic functions (Chapter 8), the use of pronouns in general will be compared to the use of full nouns and proper names. The use of different types of pronouns for pragmatic functions will also be investigated. This analysis will focus on personal pronouns as opposed to demonstrative pronouns. The remaining pronominal forms, which are relatively infrequent (§5.4.1), will be included in a category ‘other’ (see §8.2).

4.4.3 Proper names

Proper names usually refer to animate entities, such as pets or people, as in (28). Conventional names for family members, such as Mummy or Daddy, were also analyzed as proper names.

(28) Proper name (Grégoire, 2;3, French)

CHI: Ce ballon.
   ‘This balloon’

CHI: C’est pour Adrien.
   ‘It’s for Adrien’

4.4.4 Morphosyntactic coding of left or right dislocations and cleft constructions

As discussed in §2.2.3, the referential expressions studied here (nouns and determiners, pronouns, proper names) can occur in dislocations in the languages
studied. Dislocations are constructions in which an element appears on the left or right periphery of the sentence, usually resumed by a pronominal element inside the clause (De Cat, 2002). Following Lambrecht (1994: 184-188), the dislocated element is interpreted as establishing the reference and reflecting the assumed cognitive status of the referent. Therefore, dislocations were morphosyntactically coded and analyzed according to the morphosyntactic form of the dislocated element. Thus, the referential expression in (29) was coded as a personal pronoun and in (30) as a noun with a demonstrative determiner. The resumptive elements, ce in (29) and il in (30), were not coded separately.

(29) Left dislocation, morphosyntactically coded as personal pronoun (Anne, 3;0, French)

`chi:` \[ \ldots \] Lui, c(e) est un bébé.

‘He’s a baby’

(30) Right dislocation, morphosyntactically coded as a noun with a demonstrative determiner (Philippe, 3;0, French)

`chi:` Il est gros ce robot.

‘He’s big, this robot’

Following De Cat (2002), two additional criteria were used to identify dislocations\(^9\). Firstly, pronouns with non-nominative case in subject position, as in (29), were taken as an indication of a dislocation. Secondly, word order was used. If the elements of the sentence did not appear in the language’s standard word order, a dislocation was assumed. This is the case in (31) in which the subject is realized after the verb, which is in contrast with the standard word order in French: SVO. Also, the occurrence of an element between the verb and one of its arguments, as the adverbial in (32), was interpreted as an indication for a dislocated element.

(31) Right dislocation on the basis of word order and without resumptive pronoun (Philippe, 2;3, French)

`chi:` Roule bien le camion.

‘Drives well, the car’

(32) Right dislocation on the basis of word order and without resumptive pronoun (Philippe, 2;3, French)

`mot:` T’aurais pu la rencontrer.

‘You could have recognized her’

\(^9\) De Cat (2002) uses also prosodic factors as a diagnostic for dislocations. This is not possible for the French data of Philippe and Grégoire in this study, since these data do not contain information on prosody.
The last two examples also show that in child language, the resumptive element is sometimes missing, since children frequently omit elements in the early stages of acquisition. Generally, this makes it more difficult to decide what is a dislocation and what is not (De Cat, 2002). It did not, however, influence the coding or analysis in this study. The main reason to identify dislocations in this study was to avoid the double coding and double analysis of the dislocated element and the resumptive. If the dislocated element is missing, this problem did not come up. The use of dislocations for different pragmatic functions falls outside the scope of this study. Dislocations were not equally frequent across the three languages (see Notley, Van der Linden & Hulk, 2007 for a similar result). In Dutch and English, less than 1% of the referential expressions occurred in a dislocation in the child data (n=20 in Dutch and n=12 in English). In the Dutch input, 2% of the referential expressions were dislocated (n=19), whereas this construction did not occur at all in the English input. Dislocations were most frequent in French: 13% of the referential expressions in the child data were dislocated (n=676) and 11% (n=137) in the input.

Referential expressions can also occur in cleft constructions, as in (33). It has already been discussed in §2.2.3 that the presentational part, c’est in (33), was coded, since it functions as a grammatical element assigning focus (Lambrecht, 2001). Only the following referential expressions in this construction, Maminou and qui, were included in the analysis.

(33) Referential expressions in cleft construction (Anne, 3;0, French)
%act: Takes a book.

CHI: C(e) est Maminou (qui) me l’a donné là.
‘It’s Maminou who gave me that one there’

Referential expressions did not occur in it-clefts in the Dutch and English data, although this construction is grammatically possible. In French, less than 1% of all analyzable referential expressions occurred in a cleft-construction in both the child and input data.

10 Resumptive elements can also be omitted in French adult language in specific contexts (De Cat, 2002: 25).
4.4.5 Overview of morphosyntactic coding categories

Table 4.15 gives an overview of the morphosyntactic coding categories for nouns with different types of determiners and pronouns in this study. Proper names are also analyzed. In Chapter 7, the use of nouns with different types of determiners for different pragmatic functions will be analyzed in order to establish children’s sensitivity to pragmatic factors in reference. Chapter 8 focuses on the use of pronouns as opposed to the categories of nouns and proper names. The use of different types of pronouns for pragmatic functions will also be considered in Chapter 8. As was already pointed out briefly, the different determiner and pronoun types will not be analyzed with the same amount of detail: some will be combined to form a category ‘other’ in the analyses. This will be further explained in §7.2 and §8.2, where some additional methodological considerations with respect to the form-function analysis are discussed.

Table 4.15. Overview of coding categories of different types of nouns (with or without determiners) and pronouns

<table>
<thead>
<tr>
<th>Nouns and determiners</th>
<th>Pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indefinite determiner + singular noun</td>
<td>Personal pronoun</td>
</tr>
<tr>
<td>Definite determiner + noun</td>
<td>Demonstrative pronoun</td>
</tr>
<tr>
<td>Demonstrative determiner + noun</td>
<td>Possessive pronoun</td>
</tr>
<tr>
<td>Possessive determiner + noun</td>
<td>Relative pronoun</td>
</tr>
<tr>
<td>Numeral + noun</td>
<td>Reflexive pronoun</td>
</tr>
<tr>
<td>Genitive construction (D/E)</td>
<td>Numeral pronoun</td>
</tr>
<tr>
<td>Partitive determiner + noun (F)</td>
<td></td>
</tr>
<tr>
<td>Indefinite determiner + plural noun (F)</td>
<td></td>
</tr>
<tr>
<td>Grammatical bare noun</td>
<td></td>
</tr>
<tr>
<td>Grammatical bare noun as indefinite plural (D/E)</td>
<td></td>
</tr>
<tr>
<td>Ungrammatical bare noun</td>
<td></td>
</tr>
<tr>
<td>Filler + noun</td>
<td></td>
</tr>
</tbody>
</table>

Notes. D = Dutch; E = English; F = French.

4.5 Coding of the pragmatics of reference

This study investigates children’s sensitivity to the pragmatic factors of specificity, givenness in discourse and assumed familiarity of referents that are new to discourse. The pragmatic factors studied operate in seven pragmatic functions of reference in discourse (§2.3). These pragmatic functions served as the basis for the pragmatic analysis of both the input and child language. In §4.5.1, the criteria for coding person or object references...
as one of the seven different pragmatic functions will be discussed. In §4.5.2, the criteria for the pragmatic coding of left and right dislocations are considered.

### 4.5.1 Pragmatic functions

The pragmatic functions that were analyzed are outlined in Figure 4.1. The analysis of pragmatic functions was based on the surrounding discourse context in the conversations between the children and their conversational partners and also on contextual annotations. It is important to note that the subjects might not necessarily imply to convey a particular pragmatic function. A qualitative analysis of the subjects’ true intentions is, after all, impossible on the basis of production data alone. Assessment of metalinguistic awareness of discourse and discourse processes is needed for such an analysis.

![Figure 4.1. Pragmatic functions analyzed in the child and input data](image)

It is also important to note that a conversation is a process of negotiation between interlocutors (Brown & Yule, 1983: 89). Therefore, it might happen that the speaker refers to a person or object, but that the hearer does not react to the utterance, because he was not attending to the speaker. It might also occur that the hearer did not understand the utterance/reference and asks for clarification, as in (34). In both cases, the speaker will possibly repeat her utterance and reference. The referential expression will then be analyzed for a second time, according to the pragmatic function of the first referential attempt. Thus, in (34), the child’s first attempt to refer to the falling pan (*deze*) is analyzed as discourse-new, since it is a first mention. The second mention will, in this case, also be analyzed as discourse-new (and not as discourse-given), since the child’s mother asks her for clarification.
This section follows Figure 4.1 in discussing the criteria for assigning references to a particular pragmatic function.

Non-specific reference

In non-specific reference, the speaker does not refer to a particular entity, but only to a certain type of entity (§2.3). The following points were considered when analyzing a referent as non-specific: (a) whether the situation or referent talked about was hypothetical, (b) whether the referent actually existed, (c) whether the intended referent was physically present, (d) a possible generic interpretation and (e) the surrounding discourse context.

If an entity is hypothetical, reference is most likely to be only to the type of entity described, not to a particular entity. The reference to ‘water’ in (35) is therefore coded as non-specific. Additionally, if a referent does not (yet) exist or the speaker is not sure about its existence, neither the speaker nor the hearer can identify a particular entity as the referent. This was consequently coded as non-specific. In the data analyzed, this type of non-specific reference most often occurred in situations in which the child or the adult proposed to make something, as in (36).

(34) Reference after clarification questions (input to Sarah, 2:9, Dutch)

%sit: Noise of falling pots and pans.

mot: Zo.

‘Right’

chi: Au!

‘Ouch!’

chi: Doet au.

‘Hurts’

mot: Wat doet er au?

‘What hurts?’

chi: Deze doet au.

‘This one hurts’

mot: Wat zeg je?

‘What did you say?’

chi: <Deze doet au > [/] deze doet au.

‘This one hurts’

(35) Non-specific reference: hypothetical situation (Nina, 3:3, English)

%sit: Nina and her mother are playing with a toy dinner set. They play that Nina’s puppets are at a pancake restaurant.
**mot:** Are they gonna have syrup on their pancakes?

**chi:** Put some water in these glasses so that they could have water to drink.

**[36] Non-specific reference: proposal to make something (input to Matthijs, 2;3, Dutch)**

**mot:** Nou, wat ga je nou doen?
‘Tell me, what are you going to do?’

**mot:** Wil je (ee)n ketting rijgen?
‘Do you want to make a bead necklace?’

The physical context of the conversation was also used to decide between non-specific and specific reference. Referents that are physically present (see also under discourse-new-referents) were usually specific. An exception to this were situations in which the speaker desired or offered one entity out many. In example (37), the child offers his father a cookie out of a box full of cookies. This was coded as non-specific reference, since it is plausible that the child meant any of the cookies in the box, not necessarily a specific one.

In pretend play, which sometimes occurred in the data, the intended referent does not exist and is not physically present either. Pretend referents were, however, not automatically coded as non-specific reference, since in the pretend world between the conversation participants, the referent may actually be there. For instance in (38), the child uses the locative adverb daar ‘over there’ to describe the location of pretend cows, which leads to an interpretation as specific reference. In the case of pretend referents, the surrounding discourse contexts (see below) was used to decide between a non-specific or specific reading.

**[37] Non-specific reference: one out of many (Abel, 3;0, Dutch)**

**chi:** Arjen moet ook een koekje.
‘Arjen must also have a cookie’

**[38] Pretend play: specific reference on the basis of linguistic context (Matthijs, 3;0, Dutch)**

**sit:** Puppets are on the train and looking out of the window.

**inv:** Zien ze ook koeien?
‘Do they also see cows?’

**mot:** Ja, ik denk het wel hè?
‘Yeah, I think they do, right?’

**mot:** Er zijn meestal ook wel koeien in de wei.
‘Usually, there are cows in the meadows’

**chi:** Nee daar zijn koeien.
‘No, there are cows over there’
As discussed in §2.3, generic reference is included in the category of non-specific reference in this study. The speaker and hearer can only identify a class of entities, but not a particular entity or group of entities. Referents that could be interpreted generically were therefore coded as non-specific. For instance (39), the child tells about a property of rhinoceroses, that is, that they all have a horn. The referential expressions were coded as non-specific.

(39) Non-specific reference: generics (Philippe, 2;9, French)

\[\text{CHI: Les rhinoceros ils ont une corne sur la bouche.} \]

‘Rhino’s have a horn on their mouth’

The surrounding discourse context was also used to distinguish between non-specific and specific reference. If reference is non-specific, one can felicitously add the after-thought ‘but I don’t know which one of the intended class/it does not matter which one of the intended class’. For instance, in (40), the after-thought can be easily added to the underlined references, without making the utterances and references semantically odd. These references were therefore coded as non-specific. The after-thought was used as a general test to identify non-specific reference if the other, above-mentioned diagnostics (hypothetical, physical context etc.) could not be used. Alternatively, the discourse context might indicate that the referent is specific, for example because the speaker adds specific information about the referent. Also, if the speakers were talking about a past event, the references were not coded as non-specific, but as specific. For instance, in (41), the mother invites the child to talk about pancakes they had had in the zoo sometime previously. The mother therefore does not refer to pancakes in general, but the specific ones they had eaten.

(40) Non-specific reference on the basis of the discourse context (Léa and input to Léa, 2;9. French)

\[\text{GRM: Que préfères tu?} \]

‘What do you prefer?’

\[\text{GRM: Le coca?} \]

‘Coke?’

\[\text{GRM: Ou quoi d’ autre?} \]

‘Or what else?’

\[\text{CHI: Le jus de fruits (?) aussi.} \]

‘Fruit juice also’

\[\text{GRM: Le jus de fruits aussi que tu aimes?} \]

‘You also like fruit juice?’
Speakers sometimes continue to refer to non-specific referents with morphosyntactic devices that indicate a higher cognitive status than what is generally associated with non-specific reference (§2.4.1). For instance, in (42), the grandmother first refers non-specifically to ‘coke’ by means of a definite noun, le coca. In the conversation that follows, the grandmother and child continue referring to the non-specific referent ‘coke’, but now by means of pronouns (il, ce), which generally indicate a higher cognitive status than the one associated with non-specific reference. These pronominal forms still refer non-specifically, but their easy accessibility on the basis of the referent’s givenness in discourse is the more important factor in form choice, just as for specific discourse-given referents. These subsequent pronominal references were therefore not coded as non-specific reference, but as discourse-given reference (maintenance or shift, see later in this section).

(42) Non-specific reference and givenness in discourse (Léa, French, 2;9)

GRM: Tu aimes mieux le coca?
‘Do you like coke most?’

\[11\] In §2.4.1, I showed that there is a strong association between nouns with an indefinite determiner and non-specific reference. However, other forms can also be used, such as nouns with a definite determiner (see examples 40 and 42). In this respect, it is worth mentioning that in French the verbs préférer ‘prefer’ and aimer ‘like, love’ are usually followed by a noun with a definite determiner.
Finally, it is important to note that errors in expressing non-specific reference are not always clear. If in example (43), the child had used a definite determiner instead of an indefinite, a specific reading would have automatically arisen.

(43) Non-specific reference (Abel, 3;0, Dutch)

chi: Arjen moet ook een koekje.
‘Arjen must also have a cookie’

Since it is feasible that the child has one specific cookie out of many in mind, in such cases the referential expression was analyzed as specific (cf. Schafer & De Villiers, 2000 on this problem in experiments eliciting non-specific reference).

Labelling

In labelling, the speaker predicated class membership or names/identifies a specific entity (see §2.3). In this study, four diagnostics were used to decide whether or not a referent should be coded as an act of labelling. First of all, referential expressions that were answers to wh-questions from the interlocutor about the category/identity of referents that were present in the here-and-now (e.g. what’s that?/what do you see?) were analyzed as labelling. In these contexts, the adult clearly prompts the child to label referents, as in (44).

(44) Labelling after wh-question from interlocutor (Philippe, 3;3, French)

inv: Qu’est-ce que c’est ça?
‘What’s that?’

chi: C’est des fruits, des fruits.
‘That’s fruit’

[...]
Referential expressions that occurred in predicating constructions were also analyzed as labelling (see 45). Following Hickmann (2003: 66), existential constructions, such as there is/er is/il y a, were distinguished from predicating expressions. According to Hickmann, existential constructions most often serve to introduce a referent in discourse and not to identify the referent. The same applies to constructions that denote a location (e.g. there is/daar is/là, il y a). References in these constructions were only analyzed as labelling if they were preceded by a wh-question (see above).

(45) Labelling in predicating construction

a. CHI: En dat (i)s mijn speelgoed. (Abel, 2;9, Dutch)
   ‘And those are my toys’

b. CHI: That’s Pia. (Nina, 2;9, English)
   %sit: Nina points to Pia on a picture.

c. CHI: Ça c’est un gros camion rigolo. (Grégoire, 2;6, French)
   ‘That’s a big funny truck’

Especially in the early stages of language acquisition, children do not always use a full predicating construction for labelling. They may even only realize the label. Single-word utterances that refer to entities that are present in the here-and-now, and that are not expansions from a previous utterance or expanded in subsequent utterances by the same speaker, were also classified as labelling, as in (46). If the child did expand the referential expression with additional information, as in (47), the single-word utterance was not coded as labelling. According to Baker and Greenfield (1988: 20), young children often build up sequences by first uttering the new or changing information as a single-word utterance, and then use the expression as part of a following utterance to add other information to it. Following this position, the referential expression boerderij in (47) was analyzed as discourse-new reference followed by discourse-given reference (see below).
Labelling as a single-word utterance (Matthijs, 2;3, Dutch)

MOT: Ga [/] ga Evelien maar helpen.
‘Just go and help Evelien’

CHI: Kachel.
‘Stove.’

%act: Banging with hands on stove.

INV: Wat is er met de kachel?
‘What’s the matter with the stove?’

No labelling: information added in following utterance (Matthijs, 2;3, Dutch)

%sit: Matthijs wants to make a puzzle with farm animals.

MOT: Zo, nou, ga jij de puzzel maar maken dan.
‘Well, go and make the puzzle then’

CHI: Boerderij.
‘Farm’

CHI: Boerderij.
‘Farm’

CHI: Boerderij pakken. > expansion
‘Get farm’

Finally, labelling typically occurred for referents that were present in the here-and-now. It was therefore sparingly used for referents that were physically absent and only coded if the speaker explicitly provided a name for the referent in constructions as ‘that is called...’.

Specific discourse-new reference

Referents that are specific and mentioned for the first time in the current discourse, for example les mains and la poupée in (48), were analyzed as discourse-new (§2.3). The start of the transcript was taken as the start of a new conversation. If the speaker referred first to a singular entity and subsequently to a group of these entities (or vice versa), the first mentions of both the singular and plural references were coded as discourse-new. Thus, in (49), the child’s first referential expressions to the mommy (she) and the daddy (that) were analyzed separately as discourse-new. The first time that she refers to them together, by means of they, was also analyzed as discourse-new.

Discourse-new (Léa, 3;0, French)

CHI: xxx.

GRM: Comment dis tu?
‘What did you say?’
(49) Discourse-new reference to plural entity (Nina, 2;9, English)

%act: Nina starts putting all her stuffed animals to sleep on her mother’s bed.

[...]

CHI: And she’s a mommy.

CHI: And that the daddy.

CHI: I like to have the light off.

CHI: I like it dark.

MOT: Oh my # is it dark in here?

CHI: Yeah [= yes].

CHI: Cause they’re gonna sleep now.

The pragmatic function of discourse-new is further subdivided into the functions of discourse-new-mk and discourse-new-nmk (§2.3). Mutual knowledge between the conversation participants was assumed if the referent was present in the physical context of the conversation, as in (48) and (49), or if the referent was part of shared knowledge (e.g. ‘grandparents’). Furthermore, the notions of general world knowledge (e.g. ‘the queen’), uniqueness (e.g. ‘the sun’), bridging inference (e.g. ‘wheels’ of a car or ‘the cashier’ at the supermarket) were used in assessing mk between conversation participants (Clark & Marshall, 1981). Referents that were discourse-new-mk were further coded as to whether they were physically present (exophoric, exp) or physically absent (endophoric, end).

Analyzing a discourse-new referent as nmk, mk-exp or mk-end was done on the basis of the context. Both the extra-linguistic information added to the transcripts and the surrounding discourse context, such as reactions of the interlocutor, were used. For instance in (50), the mother’s reaction to the child’s reference to Dobias, a dog, indicates that the investigator is not familiar with this dog, which was subsequently coded as nmk.

(50) Discourse-new-nmk (Abel, 3;3, Dutch)

CHI: En [/] en <daar woont> [/] daar woonde Dobias ook.

‘And and Dobias lives there too’

MOT: ‘Oh’

CHI: Dobias ook.

‘Dobias also’
INV: Woont Jonas daar ook?
   ‘Does Jonas live there?’

MOT: Wacht even [>].
   ‘Wait a sec’

CHI: Dobias [<].
   ‘Dobias’

CHI: Dobias de hond.
   ‘Dobias the dog’

MOT: Ja, dan moet je even vertellen aan Gerard wie [>] dat zijn.
   ‘Yes, but then you should tell Gerard who they are’

INV: Ja [<].
   ‘Yes.’

INV: <Wie is dat dan> [>]? 
   ‘Tell me, who is he?’

MOT: <Dat weet Gerard> [<] niet.
   ‘Gerard doesn’t know that’

**Specific discourse-given reference**

References to specific entities that had already been mentioned in the discourse were coded as discourse-given, as for example the second mention of *krokodil* in (51).

(51) **Specific discourse-given (Sarah, 2;9, Dutch)**

CHI: Een krokodil!
   ‘A crocodile!’

MOT: Ojee ja.
   ‘O dearie me!’

CHI: Krokodil gaat de tas opete(n).
   ‘Crocodile is going to eat the bag’

References made by interlocutors were taken into account in this evaluation. Thus, if the child mentioned a referent for the first time, but the interlocutor had already mentioned this referent, the child’s reference was coded as discourse-given. Referents that were first mentioned in a partly unintelligible utterance were excluded (§4.3.1), but subsequent mentions of this referent in intelligible utterances were coded as discourse-given.

A distinction was made between discourse-given-maintenance and discourse-given-shift (§2.3). In referent maintenance, the subsequent reference refers to the same entity as the immediately previous reference, as in example (52), where the child continues her reference to ‘food’.
Methodology

Specific discourse-given-maintenance (Nina, 2;3, English)

CHI: That’s food.
CHI: The food fell all off.

In referent shift, the distance between the two subsequent references to the same referent is larger: there are one or more intervening references to other entities. These intervening references can either be third person references, but also first or second person references, as in (53).

Specific discourse-given-shift (Léa, 3;0, French)

%sit: Léa has just told her grandmother about her skipping rope.
GRM: Où est elle, la corde?
‘Where’s the rope?’
CHI: Elle (n’) est pas là-bas, elle est au dessus (du frigo).
‘It’s not there, it’s on top (of the fridge)’
CHI: Je vais la prendre.
‘I’ll go and get it’

It is important to keep in mind that this research only focuses on overt forms (§1.1 and §4.4). Therefore, possible instances of subject or object drop were not considered in deciding between either maintenance or shift. For example, the child’s utterance in (54) could be paraphrased as *I* screw *it*, which would make *it* a referent shift.

Specific discourse-given-maintenance (Peter, 2;0, English)

CHI: That’s engine.
CHI: Screw *it* xxx.
INV: You gonna screw *it*?

The child’s possible subject drop was, however, not taken into account in the analysis and *it* was analyzed as referent maintenance. The current research has thus been very strict by only taking overt forms into account for both the morphosyntactic and pragmatic analysis.

Pragmatically non-analyzable references

For some person or object references, it was not possible to determine their pragmatic function in discourse. These references were coded as pragmatically non-analyzable.12

12 The term pragmatically ambiguous reference will be used interchangeably with pragmatically non-analyzable reference.
The very first referential expression of a transcript was never coded for its pragmatic function, since it is unknown what preceded. The referential expression may for example be a referent maintenance or shift on the basis of a preceding reference. Similarly, if the last word(s) of the previous utterance were unintelligible, it was not possible to determine the pragmatic function of the following reference either. For instance in (55), the child’s reference to ‘Hannah’ could be either referent maintenance or shift, depending on the references in the mother’s partly unintelligible utterance. References in these contexts were therefore also coded as pragmatically non-analyzable. If the preceding utterance was fully unintelligible however, the following reference was coded as if the fully unintelligible utterance did not precede it. Coding of pragmatic functions was continued on basis of the last intelligible reference.

(55) Pragmatically non-analyzable because of preceding utterance (Matthijs, 3;0, Dutch)
%sit: Talking about going to the hairdresser’s and Matthijs’ aunt who cuts the hair of his cousins.
MOT: Wat doet Hannah dan als xxx?
‘What does Hannah do when xxx?’
CHI: Hannah (en) # Mirjam # moet haren knippen.
‘Hannah and Mirjam must cut hair’

Finally, the surrounding discourse or extra-linguistic descriptions sometimes provided too little information to assign a pragmatic interpretation to a referential expression. In (56), it is unclear if there are really ‘cows’ on the picture. It was therefore impossible to decide between specific discourse-new reference and non-specific reference to cows in general.

(56) Pragmatically non-analyzable because of too little context information (Matthijs, 2;6, Dutch)
%sit: Matthijs and the investigator are reading a book. Matthijs comments on what he sees on the picture.
CHI: Dat is # kool.
‘That’s cabbage’
INV: Is dat kool?
‘Is that cabbage?’
INV: Bloemkool?
‘Cauliflower?’
[...]
INV: Gaat de vader dat eten?
‘Is the father going to eat that?’
Pragmatically non-analyzable or ambiguous references were excluded from further analysis. The mean percentage on all person/object references were on average less than 15% in the child data and less than 10% in the input (see Chapter 6 for more details).

### 4.5.2 Pragmatic coding of left and right dislocations

The pragmatic function of referential expressions in left or right dislocations will be coded and analyzed according to the criteria described in §4.5.1. The dislocated element is considered to establish the reference (§2.2.3). Thus, in example (57), the left dislocated construction Adrien, il is as a whole analyzed as discourse-new reference. The pragmatic function of the resumptive il is not analyzed separately, since it is neither analyzed separately for morphosyntax (§4.4.4).

(57) Left dislocation, discourse-new (Grégoire, 2;6, French)

chi:  Adrien, il m’a prêté son grand cerf volant.

‘Adrien, he has lent me his big kite’

In example (58), the dislocated element son moteur and its accompanying resumptive le are separated from each other by the pronoun il, which refers to ‘robot’. This leads to two possible analyses of pragmatic function. If the dislocated structure son moteur-le were analyzed at the location of the dislocated noun, it would have the pragmatic function of discourse-given-maintenance, on the basis of the investigator’s preceding reference son moteur. If it were analyzed at the location of the resumptive, it would be reference shift from il, the robot, back to the engine.

(58) Left dislocation, discourse-given-shift (Philippe 3;0, French)

%sit:  Talking about a robot.

inv:  Et qu’est-ce qu’il, en fait de son moteur?

‘And what does he do with his engine?’

chi: Son moteur, il le laisse dans sa tête.

‘His engine, he leaves it in his head’

This type of dislocation was very infrequent: 0.5% of all referential expressions in the child data (n=18) and also 0.5% in the input (n=6). It was decided to analyze the dislocation at the location of the resumptive. The dislocated element has
been moved from the location of the resumptive to the left or right periphery of the sentence (Van der Linden & Sleeman, 2007). The sentence would also be grammatical if the dislocation were omitted. In (58), _son moteur-le_ is thus analyzed as discourse-given-shift. If the resumptive were omitted, the construction would have been analyzed at the location of the dislocation.

### 4.6 Reliability of coding

The author of this thesis coded the data. To determine the reliability of the coding scheme, a trained research assistant coded 10% of the child data independently. The mean percentage of agreement between the two coders was 81% for pragmatic functions and 98% for morphosyntactic forms.

### 4.7 Statistical analysis

The use of morphosyntactic forms (§4.4) for pragmatic functions (§4.5.1) will be investigated to establish whether children and adults are sensitive to the pragmatic factors of (a) specificity, (b) givenness in discourse and (c) familiarity. If the speakers use morphosyntactic forms differently for non-specific reference and specific references (discourse-new and discourse given), it can be argued that they are sensitive to the pragmatic factor of specificity. If they use forms differently for discourse-new and discourse-given referents, it can be argued that they are sensitive to the factor of givenness in discourse. Finally, if the speakers use forms differently for discourse-new-\_mk and discourse-new-\_nmk and/or for discourse-new-\_mk-exp and discourse-new-\_mk-end, they show sensitivity to the pragmatic factor of familiarity to the listener.

To test the relationship between two categorical variables (for example the use of morphosyntactic forms for pragmatic functions) Pearson's chi-square tests were carried out with a significance level of \(p < 0.05\). Separate age points were combined if 20% or more of the cells at individual age points had expected counts of less than five, since the statistical power of the chi-square test is drastically reduced in such cases (Agresti, 1996). The contingency coefficient \(C\) was calculated to indicate the strength of significant effects. If \(C\) is between 0.10 and 0.25, the association is weak. A moderate association is indicated in the range between 0.25 and 0.50. Associations over 0.50 are strong. A significant chi-square value indicates that the variables studied are not equally spread over the various categories, for example, that a morphosyntactic form is not used to the same extent for different pragmatic
functions. Significant chi-square values were further examined by using the adjusted standardized residual (henceforth abbreviated as asr). The asr provides information about which variables contribute to a significant chi-square value for tables that are larger than two by two. This was used in this study to examine the association or disassociation between a morphosyntactic form and pragmatic functions. The asr indicates how a particular form is used for a particular pragmatic function relative to other forms for that function and also relative to how this particular form is used for other functions. As such, it indicates if a particular form is associated with one or more pragmatic functions and/or disassociated with others. If the number of cells is large, as in this research, asr-scores between 2 and 3 are seen as major contributors to the overall (significant) chi-square value. In this research, asr-scores higher than 2 are reported.

To test the relationships between more than two categorical variables (for example the use of morphosyntactic forms for pragmatic functions by the three different Dutch children) loglinear analyses were used (Tabachnick & Fidell, 2001; Field, 2005). These can be seen as a generalization of the chi-square test. In a loglinear analysis, two- three- or four-way associations between variables are sought to explain the data. For that purpose, linear models of expected cell-frequencies are developed for the variables and the interactions between them. The first model contains all variables and the highest order interaction between these variables. The first model is always a perfect fit to the data (i.e. observed and expected frequencies are equal), since the highest-order interaction between variables also implies all lower order interactions. This first model is called the saturated model. Subsequently, as many non-significant interactions as possible are eliminated, while still maintaining an adequate fit to the data. The final model contains the fewest number of interactions, while still describing the data adequately. That is, the model does not differ significantly from the data. The chi-square value of the final model of interactions between variables is always non-significant; the relevant highest order interactions are themselves always significant. The significant interactions that are of interest to the main questions of this study are further analyzed by means of chi-square tests. Because of the large number of cases in this research, the significance level for the loglinear analyses is conservatively set at \( p < 0.01 \), to reduce the chance of finding significant effects solely due to the law of large numbers.
4.8 Outline of results chapters

In the following chapters, the results of this research will be discussed. Chapters 5 and 6 serve as background information. In Chapter 5, the use of the different morphosyntactic forms per child and adult is discussed per age point. It is necessary to establish if and when the children use the forms studied, since from that moment onwards, the investigation of this use for different pragmatic functions can start. Differences in the acquisition of morphosyntactic forms between children acquiring Dutch, English or French are also examined. Chapter 6 gives an overview of the use of different pragmatic functions in the child language and in the input. Differences will not affect the results per se, since the morphosyntactic forms are reflected as a percentage of the different pragmatic functions and not as raw numbers. It is, however, important to establish how frequent the different pragmatic functions are in adult-child discourse. This gives an indication of the children’s possibilities of expressing and practicing a particular form-function combination and also of the amount of positive evidence they receive from the input.

Chapter 5 and 6 thus serve as background to Chapters 7 and 8 where the use of forms for pragmatic functions is discussed. Chapter 7 focuses on the use of nominal forms (with and without determiners) for pragmatic functions. In Chapter 8, the use of pronouns for pragmatic functions is investigated as opposed to nouns and proper names.