アキ・コーパス

—日本語を獲得する男児の
1歳5ヶ月から3歳までの縦断観察による発話データ集—

宮田 Susanne

The Aki Corpus: Longitudinal Speech Data of a Japanese Boy aged 1;5.7〜3;0.0

Susanne Miyata

はじめに

アキ・コーパスは日本語を獲得する男子の言語発達を見つめた縦断的観察によるデータである。アキ君と母親との会話を56セッションにわたって（合わせて約48時間）ビデオ録画し、JCHAT フォーマットで書き留めた。アキ・コーパスは95年3月に CHILDES データベースに含まれ、公開されて以来、anonymous ftp で自由にアクセスできるようになった。

1．CHILDES と JCHAT v. 1.0

CHILDES（Child Language Data Exchange System）のデータベースは Catherine Snow（Harvard University）と Brian McWhinney（Carnegie Mellon University）を中心に1984年に作られ、現在（95/3）19言語での91コーパスを含む。

しかし、今までは日本語のデータが、日本語・デンマーク語のバイリンガルデータ（ハヤシ・コーパス）を除いて、CHILDES に入力されていない。その理由の一つとして、CHILDESが最近までローマ字使用（ASCII コード）に限られ、漢字かな混じりのデータを扱えなかったことである。その上、日本語のローマ字表記法は統一せず（特にヘボン式・訓令式・日本式表記法の共存、そして分かり書きの問題）データのローマ字化をさらに困難にしたように考えられる。

そこで、1994年に日本語 CHILDES 計画が作られ、日本語発話標準書式（JCHAT v. 1.0）が
できた（Oshima–Takane・Mac Whinney 編 [in press]）。アキ・コーパスはこの JCHAT フォーマットに従っている。

2. アキ・コーパスの概要

2.1. History

This data was collected during the preparation of my dissertation about question acquisition. Aki was (after Re and Nao) the third child observed longitudinally. For Aki’s observation I applied the same method used for the observation of the other two children: that is, once a week for about one hour at his home while playing with his mother.

In the previous observations it had proved convenient for both mother and observer, to fix weekday and time. As Aki was quite a late riser, we decided to start each session about 10 o’clock in the morning. After a short period of excitement, the child would settle down to play. The video recordings started usually about 10:20. For the recording I held the camera in my lap (rather than in front of my face), a method that had proved effective in prior observations. The setting was free indoor play. The mother was instructed to ‘make the child speak’, but there were no regulations concerning the kind of play.

This data was computerized from the beginning, however in a slightly different format than JCHAT (data was organized in colons rather than in lines. Comments were written in German). The input was done directly from the video tape without handwritter notes. After the file was completed, the file was counterchecked once more against the video tape.

The transcription was done in Romaji (Hebon) rather than in Japanese script, in order to better preserve the actual pronunciation. I also used UNIBET symbols, especially when the meaning of the utterance was unclear. For slightly deviant items with clear meaning no phonetical transcription is provided. The transcription was done in JCHAT 1.0 Hebon, using Morikawa-shiki for morphemecization.

Situational cues were provided to a certain extent, to make it possible to follow the conversation without visual cues.

2.2. Codes

Concerning the intonation of questions I used $\text{FIN}$ (falling intonation) and $\text{RIN}$ (rising intonation). Where unmarked, assume rising intonation.
Concerning the classification of wa—Questions (see Miyata 1992, 1993), I used the following four codes:

<table>
<thead>
<tr>
<th>Code</th>
<th>Japanese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>$WAP$</td>
<td>wa—Question (Place)</td>
<td>papa wa ?</td>
</tr>
<tr>
<td>$WAN$</td>
<td>wa—Question (Name)</td>
<td>kore wa ?</td>
</tr>
<tr>
<td>$WAE$</td>
<td>wa—Question (Educational)</td>
<td>gomen ne wa ?</td>
</tr>
<tr>
<td>$WAG$</td>
<td>wa—Question (General)</td>
<td>papa wa ookii. mama wa ?</td>
</tr>
</tbody>
</table>

2.3. Biographical data

Aki was born on 27. Sep. 1987 in Nagoya, the firstborn child. His mother was 31 years old at the time of his birth. Pregnancy and delivery were normal. Aki’s birth weight was 2870g. His physical development was normal, aside from a 6-day hospital stay (2;4.30–2;5.4) due to a small operation (surgical cut of a short thumb sinew), and he was healthy throughout the observation.

Aki was an active, curious, fearless child, very interested in books and stories. However, his concentration span was quite short, and he would soon grow weary. His pronunciation was very clear. He uttered his first word at 1;8. At present (Feb. 1995), he is an average student in the 1st grade of primary school.

Participants:

AMO, Mother, called “Okaasan”, 32 years, pianist, part-time lecturer in the piano section of a senior highschool in Nagoya, and gives private lessons, and concerts.

AFA, Father, called “Otoosan”, associate professor for biogenetics at a University in a nearby town to Nagoya

REE, 2-year-old younger brother Ree, called “Reechan”, born 22–AUG–89 (Aki’s age: 1;10:26)

OBA, baby sitter, called “Obasan”, 61 years, no university degree

BAA, Grandmother, maternal, called “Baaba”, former primary school teacher

OOB, Grandmother, paternal, called “Obaasan”, housewife

SUZ, Investigator, called “Suuze(san)”, friend of AMO

AKI, AMO, AFA, REE live together. Occasionally BAA and sometimes also OOB come to visit.
2.4 Pseudonyms

Aki's parents gave their kind consent for the publication of this data.

Although they consented to the use of their actual names, I have decided to replace all last names (except my own) throughout the corpus in order to preserve a minimum of privacy.

2.5 Table of Contents

<table>
<thead>
<tr>
<th>file name</th>
<th>old name</th>
<th>age</th>
<th>date</th>
<th>partic. except AKI, MOT, SUZ</th>
<th>length</th>
<th>numbers of utt.</th>
<th>mlu</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKI01 AC</td>
<td>1:5.7</td>
<td>6—MAR—89</td>
<td>n(one)</td>
<td>0:11:20</td>
<td>51</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>AKI02 AD</td>
<td>1:6.10</td>
<td>6—MAR—89</td>
<td>n</td>
<td>0:25:15</td>
<td>90</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>AKI03 AE</td>
<td>1:7:14</td>
<td>1—MAY—89</td>
<td>n</td>
<td>0:13:35</td>
<td>80</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>AKI04 AF</td>
<td>1:8:23</td>
<td>19—JUN—89</td>
<td>n</td>
<td>0:37:20</td>
<td>179</td>
<td>1.016</td>
<td></td>
</tr>
<tr>
<td>AKI05 AG</td>
<td>1:9:20</td>
<td>17—JUL—89</td>
<td>n</td>
<td>0:25:30</td>
<td>173</td>
<td>1.021</td>
<td></td>
</tr>
<tr>
<td>AKI06 AH</td>
<td>1:10:0</td>
<td>27—JUL—89</td>
<td>n</td>
<td>0:34:30</td>
<td>300</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>AKI07 AI</td>
<td>1:11:29</td>
<td>25—SEP—89</td>
<td>REE AFA</td>
<td>0:21:30</td>
<td>111</td>
<td>1.049</td>
<td></td>
</tr>
<tr>
<td>AKI08 BA</td>
<td>2:0:15</td>
<td>2—OCT—89</td>
<td>REE OBA</td>
<td>0:38:50</td>
<td>211</td>
<td>1.048</td>
<td></td>
</tr>
<tr>
<td>AKI09 BB</td>
<td>2:0:12</td>
<td>9—OCT—89</td>
<td>REE OBA</td>
<td>0:35:30</td>
<td>227</td>
<td>1.058</td>
<td></td>
</tr>
<tr>
<td>AKI10 BA</td>
<td>2:0:19</td>
<td>16—OCT—89</td>
<td>n</td>
<td>0:36:30</td>
<td>210</td>
<td>1.076</td>
<td></td>
</tr>
<tr>
<td>AKI11 BD</td>
<td>2:0:26</td>
<td>23—OCT—89</td>
<td>OBA</td>
<td>0:38:00</td>
<td>178</td>
<td>1.034</td>
<td></td>
</tr>
<tr>
<td>AKI12 BE</td>
<td>2:1:3</td>
<td>30—OCT—89</td>
<td>REE</td>
<td>0:47:15</td>
<td>218</td>
<td>1.110</td>
<td></td>
</tr>
<tr>
<td>AKI13 BF</td>
<td>2:1:10</td>
<td>6—NOV—89</td>
<td>n</td>
<td>0:47:35</td>
<td>209</td>
<td>1.029</td>
<td></td>
</tr>
<tr>
<td>AKI14 BG</td>
<td>2:1:17</td>
<td>13—NOV—89</td>
<td>REE</td>
<td>0:36:10</td>
<td>296</td>
<td>1.226</td>
<td></td>
</tr>
<tr>
<td>AKI15 BH</td>
<td>2:1:24</td>
<td>20—NOV—89</td>
<td>REE BAA</td>
<td>0:50:50</td>
<td>378</td>
<td>1.101</td>
<td></td>
</tr>
<tr>
<td>AKI16 BI</td>
<td>2:2:0</td>
<td>27—NOV—89</td>
<td>REE BAA</td>
<td>0:48:00</td>
<td>282</td>
<td>1.166</td>
<td></td>
</tr>
<tr>
<td>AKI17 CA</td>
<td>2:2:11</td>
<td>8—DEC—89</td>
<td>n</td>
<td>0:49:30</td>
<td>396</td>
<td>1.155</td>
<td></td>
</tr>
<tr>
<td>AKI18 CB</td>
<td>2:1:4</td>
<td>11—DEC—89</td>
<td>REE OBA</td>
<td>0:46:40</td>
<td>369</td>
<td>1.209</td>
<td></td>
</tr>
<tr>
<td>AKI19 CC</td>
<td>2:2:122</td>
<td>18—DEC—89</td>
<td>REE</td>
<td>0:48:40</td>
<td>383</td>
<td>1.352</td>
<td></td>
</tr>
<tr>
<td>AKI20 CD</td>
<td>2:3:0</td>
<td>27—DEC—89</td>
<td>REE</td>
<td>0:48:25</td>
<td>208</td>
<td>1.438</td>
<td></td>
</tr>
<tr>
<td>AKI21 CE</td>
<td>2:3:4</td>
<td>31—DEC—89</td>
<td>REE AFA</td>
<td>1:02:20</td>
<td>546</td>
<td>1.378</td>
<td></td>
</tr>
<tr>
<td>AKI22 CF</td>
<td>2:3:12</td>
<td>8—JAN—90</td>
<td>REE</td>
<td>1:00:00</td>
<td>526</td>
<td>1.588</td>
<td></td>
</tr>
<tr>
<td>AKI23 CG</td>
<td>2:3:18</td>
<td>17—JAN—90</td>
<td>REE</td>
<td>1:00:00</td>
<td>594</td>
<td>1.807</td>
<td></td>
</tr>
<tr>
<td>AKI24 CH</td>
<td>2:3:26</td>
<td>22—JAN—90</td>
<td>REE</td>
<td>1:00:00</td>
<td>447</td>
<td>1.987</td>
<td></td>
</tr>
<tr>
<td>AKI25 CI</td>
<td>2:4:4</td>
<td>31—JAN—90</td>
<td>REE</td>
<td>1:00:00</td>
<td>520</td>
<td>1.933</td>
<td></td>
</tr>
<tr>
<td>AKI26 DA</td>
<td>2:4:9</td>
<td>5—FEB—90</td>
<td>REE</td>
<td>1:00:00</td>
<td>436</td>
<td>1.667</td>
<td></td>
</tr>
<tr>
<td>AKI27 DB</td>
<td>2:4:18</td>
<td>14—FEB—90</td>
<td>REE</td>
<td>1:00:00</td>
<td>428</td>
<td>1.759</td>
<td></td>
</tr>
</tbody>
</table>
2.6. Warnings

a) This data is NOT suitable for the study of the mother’s overall language behavior, except for questioning and answering behaviour.

The Aki Data was originally sampled for the study of question development (cf. 2. 1. History). For this purpose, I have recorded all utterances of the child, and also the mother’s questions and answers to questions, but have cut repetitions and longer narrations of the mother. These
cut parts are unfortunately "unmarked". In preparing the data for CHILDES, I considered going through the original tapes and add the missing parts of mother's utterances to the data files, but after a consideration of the pros and cons, I decided to publish the Aki Data in this incomplete form.

b) Reliability was not checked.

c) The length of the observational sessions differ.

Due to my research interest, the child's condition, and sometimes also due to technical problems, the length of the sessions, especially in the first year of observation, were sometimes less than 60 min. The actual length is noted in the Warning Header of each file, as well as in 2.5. Table of Contents.

3. ファイル・サンプル

@Begin
@Participants: AKI Akifumi Target child, AMO Okaasan Mother, REE Ree Brother, SUZ Suuze Investigator,
@Filename: AK156 OLD: GD30101
@Age of AKI: 3:0:0
@Date: 27–SEP–1990
@Coding: JCHAT 1.0 Hebon 95/1
@Warning: recorded time: 1:00:00
@Comment: AKI: Number of utterances 426, morphemes 1127 mlu 2.646, sd 1.870

⋯⋯[省略]⋯⋯
*AMO: ja, kondo kore ni shiyoo ka?
%cod: $Q \text{FIN}$
%exp: picture of car made of board and 4 wheels
*AKI: un.
*AMO: jidoosha ga dekiru yo.
*AMO: koko de mottete.
*AKI: a, kotchi.
%act: fixes first wheel
*AMO: dekita ne.
*AKI: kotchi wa?
%cod: $Q \text{WAG}$
%act:  takes second wheel, holds it onto the first one
*AMO:  sotchi moo ii yo., kore de.
*AMO:  ja, kondo kotch shiyoo ka?
%cod:  $Q $FIN
*AKI:  un.
*AMO:  taiya totte.
*AMO:  Ree-chan kita.
*AKI:  kore.
*AMO:  n?
%cod:  $Q
*AKI:  kore ga ii yo.
*AMO:  moo chotto +...
*AKI:  kore de +...
%act:  takes long pin
*AKI:  motto nagai no wa?
%cod:  $Q $SWAP
*AMO:  n?
%cod:  $Q
*AMO:  kanazuchi kashite.
*AMO:  moo ikko atta ne, nagai no.
*SUZ:  Aki-chan no ushiro.
*AMO:  hai, daiku-san shiyoo?
%cod:  $Q
%tim:  0:18:45
*AMO:  hai, Aki-chan yatte goran.
*AKI:  O#kaa-san yatte [: yaru-te].
*AMO:  hai.
*AMO:  koko tataku no ne.
*AKI:  un.
*AMO:  koo yatte tataku to nakanaka haitte+ikanai kara(because) ne.
%cod:  $NEG
*AKI:  un.
*AMO:  nani noseru?
%cod:  $Q
%exp:  all four wheels fixed, so car can transport something
*AMO: dekita yo.
*AMO: nani noseru?
%cod: $Q
*AKI: toottara [tooru−tara] dame.
*AKI: a ima tootchatta [tooru−chau−ta].
*AMO: chotto mane shita tsukutte + miyoo ka?
%cod: $Q
%act: starts building after model picture
*AKI: ima fumikiri to tsukutteru [tsukuru−teru] n da kara (because).
*AMO: fumikiri?
%cod: $Q
*AKI: kowashitara [kowasu−tara] dame!
%add: REE
*AKI: kowashitara [kowasu−tara].
%add: REE
*AKI: kowashitara [kowasu−tara]!
%add: REE
*AMO: shinai + shinai.
%cod: $NEG
*AKI: kowashitara [kowasu−tara] dame.
%cod: $NEG
*AMO: shinai yo.
%cod: $NEG
*AMO: a. gomen + gomen!
%exp: Ree destroys building
……[省略]……
@End

文献・連絡先

Mac Whinney, Brian 1995 (2nd ed. i. p.)
The CHILDES project: tools for analyzing talk
Hillsdale, NJ (Lawrence Erlbaum Assoc.)

宮田 Susanne 1993
Japanische Kinderfragen
Zum Erwerb von Form − Inhalt − Funktion von Frageausdrücke
Hamburg (OAG)
宮田 Susanne 1992
ダレは who より遅い?
—日本語を獲得する子どもとその母親的人物質問の使用について—
愛知淑徳短期大学研究雑要第30号，1991/11，209－219

宮田 Susanne 1992
「パパワ」—子どもの「ワ」を含む質問について—
愛知淑徳短期大学研究雑要第31号，1992，151－155

宮田 Susanne 1994
MinJ—CHAT 0.9 Hebon (Preliminary Version)
—CHILDES 用日本語データ入力フォーマット—
愛知淑徳短期大学研究雑要第33号，1994/6，153－166

Ono, Seiko Fukuchi 1994
A Corpus of a Japanese Girl: ARISA (1;6.2): Part 1
愛知淑徳短期大学研究雑要第33号，1994/6，81－108

大島・高根百合子，マックウィンニー，プライアン（編）1995（i. p.）
日本語のためのCHILDES マニュアル

寺尾康1989
発話資料のデータベース化
言語1989/6，122－124

CHILDES プロジェクト連絡先：
Child Language Data Exchange System
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15132 U. S. A.
childes@andrew.cmu.edu (Brian MacWhinney)

JCHAT コアプロジェクト連絡先：
日本国外：yuriko@hebb.psych.mcgill.ca (Yuriko Oshima–Takane)
日本国内：h44816g@nucc.cc.nagoya-u.ac.jp or GCB02445 (Susanne Miyata)

後 記
このコーパスの作成にあたって，アキ君とアキ君のご家族に長いあいだに協力していただき，心からお礼を申し上げます。データの文字化にあたって，JCHAT メンバー，特に中嶋夫さん（大阪大学），大野裕美さん（愛知淑徳短期大学），森川・ポール裕美さん（University of Kansas），大島・高根百合子さん（McGill University），Brian MacWhinney さん（Carnegie Mellon University），そして Beverley Curran（愛知淑徳短期大学）に貴重なコメントをいただきました。ここに記して感謝いたします。