IV Corpus linguistics

Susanne Miyata and Brian MacWhinney

10 CHILDES for Japanese: Corpora, programs, perspectives

1 What is CHILDES

Apart from the invention of audio recording technology, few developments have had a greater impact on language research than the introduction of affordable microcomputers and the advent of the Internet in the seventies. These developments opened up increasingly powerful methods for storage, analysis, and sharing of large amounts of speech data. One of the most influential linguistic projects in this area has been the Child Language Data Exchange System (CHILDES: MacWhinney 2000). Founded in 1984, CHILDES is now a huge database of over 60 million words containing speech data of thousands of children and their caretakers from 34 languages, including Japanese. In 2001, the CHILDES framework was extended to include adult spoken language data as a part of the larger TalkBank Project. TalkBank (http://talkbank.org) and CHILDES (http://childes.talkbank.org) now include L2 acquisition data, conversation analysis data, data on phonological development, and a wide variety of clinical data.

One of the reasons for the international success of CHILDES has been its emphasis on the precise control of the format of the transcriptions, not only across corpora but also across languages. The CHAT transcription format (an acronym for Codes for the Human Analysis of Transcripts) consists of a small set of obligatory basic rules and a large variety of optional rules and codes for detailed coding of spoken patterns in terms of phonetics, morphosyntax, and Conversation Analysis. All CHILDES and TalkBank corpora conform tightly to this common CHAT standard, which can be validated through precise XML definitions.

Because the corpora adhere to this strict format, it is possible to construct accurate computational analyses of many facets of speech. The Computerized Language Analysis program (CLAN: MacWhinney, 2000) provides a range of analysis and utility commands fit especially to language analysis. Search commands like FREQ (computing the frequency of all or selected items), KWAL (producing lists of utterances containing a specified word or morpheme), COMBO (searching combinations of two specified words or morphemes), and COOCCUR (looking for patterns of co-occurrence between words) cover the basic needs. These commands can be run on multiple files, and adjusted with option switches specifying the speaker, the sample size, the data to be analyzed, the output format, and so forth. These option switches have been developed over the years in response to specific requests from child language researchers, and they cover basically all aspects of linguistic corpus analysis.

(Unicode 9 15/8/15 21:05) WDG-New (170mm×240mm) DGMetaSerifScience (OpenType) 0000 pp. 253–282 1666 Minami_10_Miyata (p. 255)

CLAN also includes a number of more specialized programs, such as CHAINS (tracking sequences of interactional codes), CHIP (examining repetitions and expansions between two speakers), DIST (examining patterns of separation between speech act codes), FREQPOS (tracking frequencies in various utterance positions), PHONFREQ (computing the frequency of phonemes in various positions), or MODREP (matching the child's phonology to the parental model). Major developmental assessment scores like the Mean Length of Utterance (Brown 1973) or the Developmental Sentence Scoring (Lee 1974) can be obtained in a single automatic step with the MLU and DSS commands.

For research questions that go beyond word-based configurations, such as speech acts, code switching, or speech errors, the use of coding tags has proved useful. In CLAN's "Coder Mode", tags can be attached to utterances by selecting from a set of options. The tags themselves are derived from a library file called codes.cut. This file contains a coding scheme that can be freely modified according to the researcher's aims. Once the transcriptions are coded, the coding tags can be searched and analyzed like ordinary words. Other types of codes can be inserted automatically. Examples of this type are the morphological and syntactic tags inserted by the MOR and GRASP commands. Both commands are based on language-specific library files. As grammatical tags provide a good basis for any kind of grammar-oriented research, most CHILDES corpora are now tagged for morphological information.

Other coding systems focus on speech errors, speech acts, and conversational features. Errors can be simply marked by [*], but it is possible to mark errors as involving phonology [* p], semantic [* s], neologism [* n], dysfluency [* d], or morphology [* m]; sentence level errors can be marked by postcodes (e.g., [+ gram]). Each error type can be further coded using extensions such as [* n:uk] for a neologism with an unknown word target, or they can be characterized further on the %err tier.

Speech act coding based on the Inventory of Communicative Acts – Abridged (INCA-A) proposed by Ninio et al. (1994) can be performed semi-automatically with a utility called Coder's Editor. On a separate %spa tier, codes for the type of interchange are combined with illocutionary force codes, as in the following example (1) in which "dhs" stands for "discussing hearer's sentiment", "yq" indicates a yes/no question.

(1) *MOT: are ☺ you oka:y ☺ ↗ ?
%spa: $dhs:yq

This example also illustrates the use of special symbols for features of Conversation Analysis (CA) coding. Specifically, the funny faces indicate the beginning and end of "smile voice", and the up pointing arrow indicates final rising contour. Through these and other features, CHAT supports all of the codes of traditional Jeffersonian CA analysis. In addition, the use of special marks for the beginning and end of overlaps allows the INDENT program to automatically realign overlaps. Using

these and other methods, the CHAT format allows the combination of CA coding with conventional orthographic and speech act coding, thus making possible a wide range of analytic possibilities.

CLAN also provides extensive support for the linkage of individual words or utterances to audio or video media. A double click on an utterance launches the media source, and the corresponding video or audio is played back. The process of linking the video or audio to the utterances is facilitated by utility commands and specific linkage modes ("Sonic Mode" and "Transcriber Mode"). This support for fine-grained temporal linkage has greatly enhanced the quality and reliability of data transcription. At any moment in the process of transcribing and analyzing data, it is possible with a double click to go back precisely to the original utterance, and to correct the transcription, if necessary.

In addition to the analysis commands, a number of utility commands facilitate transcription, tagging, coding, and checking of the data, and translate external formats like ELAN, EXMARaLDA, Phon, Praat, and Transcriber to and from CHAT format.

The utility commands and transcription modes provide aids that greatly facilitate the transcription process. Nevertheless, the basic task of transcribing utterances has not been automated at present, and still constitutes a considerable workload, impeding any rapid expansion of the Japanese section of the database. In the following, we will look at the development of CHILDES in Japan, including the specific difficulties deriving from the fact that Japanese uses a non-Latin script, and try to analyze the problems and chances of the current situation.

2 The start of CHILDES in Japan: Romanization format and morphological tagging

Compared to English, Japanese research based on speech data corpora had a late start. The main reason for this can be found in the ideographic script, and the problems it posed for word processing. While Latin script typewriters became common already in the early 20th century, comparable machines were not commonly available for the complex Japanese script. As a result, research activities relied on handwriting long into the late 1980s. For example, the National Institute for Japanese Language and Linguistics published the longitudinal speech data of the boy Taro (1;0–3;11) in 1983 in the form of a facsimile of handwritten notes (Kokuritsu Kokugo Kenkyūjo 1981–1983). When the first electronic word processors appeared in the 1980s, it became possible to type and print a high number of ideograms, but the storage capacity was low, and the format was not compatible between the machines of different manufacturers. However, with the advent of the Japanese input method editor ATOK (JustSystems) in combination with the first Japanese word processing

software ("Ichitaro") in 1985, microcomputers gradually started to become popular, and a more effective way of collecting, processing, and sharing of Japanese speech corpora became possible.

2.1 Romanization of the Japanese script, and the definition of morpheme and word boundaries

In the early 1990s, when the first rumors about CHILDES spread among Japanese linguists, several researchers started to enter their child speech data in the CHAT transcription format. But they soon faced a number of difficulties related to the Japanese script. At that time, the CHILDES data files were limited to ASCII code, a basic set containing the 26 alphabetic letters, a handful of punctuation marks, the numbers 0 to 9 and some control characters. This meant that the transcription of Japanese data had to be performed in the Latin alphabet. There are two traditional ways of Romanization for Japanese (the Kunrei-style and the Hepburn-style), but it turned out that neither one was stringent enough for database purposes.

The Kunrei-style Romanization is grounded on the Kana syllabary matrix developed in the 11th century, which is arranged according to the first consonant in the order of *a i u e o*. The first line includes the syllables consisting only of a vowel (*a i u e o*), the second line includes the syllables starting with *k* (*ka ki ku ke ko*), and the following lines proceed with *s*, *t*, *n*, *h*, *m*, *y*, *r*, and *w* as first consonant. New pronunciations that became necessary for the pronunciation of Chinese loanwords imported in the following centuries were expressed with combinations of small Kana (e.g., /tʃa/ is written with the syllable ち *ti* with a small や *ya* following: ちや). In the thousand years since this system was established, the pronunciation of some syllables has changed (Coulmas 1989). Here the difference between the two Romanization methods originates. The Kunrei-style presents the historical perspective, and transcribes the syllables according to their position in the syllabary (e.g., *sa si su se so*; *ta ti tu te to*). The Hepburn-style, on the other hand, reflects the modern pronunciation (e.g., *sa shi su se so*; *ta chi tsu te to*). Similarly, the combinations with small Kana are transcribed as *tya* in the Kunrei-style, but as *cha* in Hepburn-style. Traditionally, the Kunrei-style is used in linguistics and grammar theory, and the Hepburn-style in second language teaching and research. In daily life, the Hepburn-style is preferred for transcriptions for the use by foreigners, such as road signs, but, in practice, the two systems are not strictly separated, and often a mix of both is used. For database purposes this meant that a number of decisions concerning the Latin orthography of Japanese had to be made.

A second problem was that Japanese Kana-Kanji script does not mark word boundaries, and, as a result, there is no commonly shared consensus about what constitutes a "word". The definition of words affects the results of most linguistic analyses since words are one of the basic analysis units of the CHAT format. Soon

discussions among researchers started, and when Yuriko Oshima-Takane called for participation in 1993, the JCHAT Project came into being. In 1995, the first JCHAT Workshop was held, and the first Japanese CHILDES manual was published (Oshima-Takane and MacWhinney 1995). This manual included an introduction to the CHAT format and the CLAN program, but the main focus was on the Romanization and the definitions of word classes and morphemes. Oshima-Takane and her colleagues formulated conclusive rules for both Romanization systems, the Kunrei-style and the Hepburn-style. However, in the following years it became clear that it was difficult to maintain support for two systems. Because the language corpora being collected were all formatted in the Hepburn-style, program development concentrated on the Hepburn-style, and currently analysis programs in the Kunrei-style are not supported.

The second focus of the JCHAT manual was on the definition of word boundaries and morphemes. Case and final particles were transcribed as separated words, while inflections and pre- and suffixes were attached to the stem. An utterance like Example 2a looks in Japanese CHAT format like Example 2b. The speaker is indicated by a 3-letter code preceded by *, and the utterance follows in Latin script.

(2a) *Kirin-san ga o-hiru-gohan tabe-ta ne.*
giraffe-HON NOM HON-noon-meal eat-PST TAG
'Mr. Giraffe has eaten lunch, right?'

(2b) *MOT: kirinsan ga ohirugohan tabeta ne. [CHAT format]

2.2 The development of the grammar analysis tools for Japanese

2.2.1 JMOR and MLU

The following years were devoted to the elaboration of word and morpheme definitions and the development of grammar analysis tools. The second edition of the JCHAT Manual (Oshima-Takane et al. 1998) already included a chapter on JMOR, the Japanese version of the morphological tagging command MOR. JMOR (Naka and Miyata 1999) produces a separate %mor tier for morphological tagging with information on word class, stem, and inflections for each word. Simultaneously, the first author of this chapter developed the closely connected Wakachi98 (Miyata and Naka 1998; Miyata 2003), a guideline for the segmentation of words and the definition of word classes and affixes based on Masuoka and Takubo (1995). The need to define every single word encountered in the mother-child conversations of the database forced us to make decisions also for grey zone cases that are usually not accounted for in grammar analysis, and led to a comprehensive MOR lexicon with currently ca. 12,000 entries including more than 300 prefixes and suffixes (Miyata

and Naka 2010). The current version JMOR05 includes English translation. Both JMOR and Wakachi2002 can be downloaded separately from the CHILDES homepage at <http://childes.talkbank.org/morgrams/>. Example 3 gives the MOR output for the sample sentence used above (Example 2).

(3) *MOT: kirinsan ga ohirugohan tabeta ne.
%mor: n|kirin-san=giraffe_HON ptl:case|ga=NOM
o#n|+n|hiru+n|gohan= HON_lunch
v:v|tabe-PAST=eat ptl:fina|ne=TAG.

The MOR command works semi-automatically in combination with the POST command. MOR produces all possible analyses of each word. For example, the nominative case particle *ga* is analyzed as conjunctional particle (ptl:conj|ga=although), case particle (ptl:case|ga=NOM), and noun (n|ga=moth). The POST command chooses the best-matching solution according to the context. In example above, the conjunctional particle and the noun are both ruled out, because *ga* follows a noun. POST uses statistical information obtained by training based on a sample corpus. The accuracy rate is rather high (for the Tai corpus we obtained an average of 0.8 mistakes in 100 utterances), although the error rate might be higher for other corpora. The disambiguation of homophones of the same word class needs a check of the context (鼻 n|hana=nose versus 花 n|hana=flower; 掛ける v:v|kake-PRES=hook versus 書ける v:c|kak-POT-PRES=write), and cannot be performed automatically, but has to done by hand. The "disambiguation mode" facilitates this procedure.

Once the morphological tags are attached and checked for reliability, they can be used in multiple ways. Because the %mor line contains grammatical information on word class, stem, and eventual suffixes, precise database searches on word classes or specific grammatical inflections become possible. For example, the frequency of certain word types or inflections (FREQ), lists of utterances containing these word types or inflections (KWAL), as well as combinations (COMBO) and patterns of co-occurrence (COOCCUR) can be easily computed. Also cross-linguistic semantic analysis is possible with the help of translation tags: semantic groups like color terms or mental verbs can be searched simultaneously in a range of languages with a simple search file containing a list translation terms for colors (*=red, *=blue, etc.) or mental verbs, respectively (*=think, *=believe, etc.).

Moreover, the %mor line serves as the starting point for the automatic computation of developmental assessment, such as the Mean Length of Utterance (MLU: Brown 1973), a widespread measure for the grammatical development up to age four. Although the idea of measuring the average sentence length had appeared as early as 1943 in the Japanese language acquisition research literature (Ushijima and Moriwaki 1943), it was not further utilized except by Murata (1968), probably because in Japanese MLU is not as easily computed as in English, due to the continuous Japanese script, which makes a word count less obvious than for English.

Murata (1968) presents longitudinal MLU data of five children between 1;0 and 1;11, but does not provide any definition of the words (*go*) counted. In the 1990s several proposals for a Japanese MLU were discussed (Miyata 1999, 2012a; Ogura et al. 1997; Watamaki 1981, 1993). All of these versions can be performed by CLAN; however, none of them is standardized yet. The currently available results suggest that the average Japanese MLU curve resembles strongly its English counterpart.

2.2.2 DSSJ

Besides MLU, the Developmental Sentence Score (DSS) proposed by Lee (1974) can be computed with CLAN. The DSS command works with a set of language-specific scoring rules distributed with CLAN. DSS scores certain grammatical morphemes of eight grammatical subareas ('indefinite pronouns and noun modifiers', 'personal pronouns', 'main verbs', 'secondary verbs', 'negatives', 'conjunctions', 'interrogative reversals', and 'WH-questions') according to their complexity and average age of acquisition, thus measuring grammar development more directly than MLU. This division into eight areas makes it possible to extract detailed scores for grammatical subareas such as verb inflection or pronouns.

The Developmental Sentence Score for Japanese follows basically the same idea (DSSJ: Miyata et al. 2006, 2009, 2013, Miyata, Nisisawa, and Otomo 2005). But, as the language structure of English and Japanese is different in many aspects, a completely new set of grammatical areas and representative morphemes had to be developed. Using updated CHILDES corpora (Aki, Ryo, Tai, Arika, Nanami, Tomito, Asato, as well as the data from Taro (Kokuritsu Kokugo Kenkyūjo, 1982a, b) the first author and her colleagues selected grammatical items that were acquired in the same order for the eight children. The selected items were grouped in the following nine subareas: 1) verb last inflection, 2) verb middle inflection, 3) copula inflection, 4) conjunctions and conjunctive particles, 5) noun phrase structure and compounds, 6) case, topic, focus, and quotative particles, 7) adverbs, 8) sentence modality markers and formal nouns, and 9) final particles. Each subarea was divided into five developmental stages. For example, in the area of case particles, the nominative *ga* and locative *ni* were grouped into stage 1, while the accusative particle *o* was acquired later and therefore grouped into stage 3.

The resulting DSSJ Scoring Table was confirmed with a cross-sectional sample consisting of adult-child interactions of 84 children (31 boys and 53 girls) including six age groups ranging from 2;8 to 5;2, each group consisting of 14 children (Miyata et al. 2009, 2013). For each child, 100 utterances were analyzed. The high correlations of the children's overall DSSJ score with MLU and age in months indicated a high reliability of DSSJ, also in the higher age range. We also investigated in which subareas high-scoring children would excel, and it turned out that most

high-scoring children (defined as scoring higher than 1 SD than the mean DSSJ overall score) did not score highly in all subareas equally, but scored highly in various but different subareas. For the younger children between 2;8 and 3;8 there was a tendency to score high on finite verb inflection, conjunctions, and case particles, while most older children between 4;2 and 5;2 achieved high scores with conjunctions and noun phrases including compounds. This reflects the developmental tendency to progress from simple sentences with finite verbs and particles to complex sentence constructions using conjunctions and complex noun phrases.

This first evaluative study suggests that DSSJ may be valuable for studying grammatical development in both typical and non-typical children. A first comparison of the language development of children with pervasive developmental disorders and children with severe mental retardation with typically developing children showed that children with pervasive developmental disorders tend to score considerably lower than the typically developing and also the mentally retarded children (Miyata, Otomo, and Nisisawa 2007).

A recent development in CHILDES is the syntax tagger GRASP (Sagae et al. 2010), which computes grammatical dependency relations on the basis of the %mor tier. A Japanese version of GRASP has been released in 2010 (Miyata, Sagae, and MacWhinney 2013). The necessary library files are included in the JMOR folder mentioned above.

GRASP provides information on dependency relations in a sentence and also the case role of each arguments. Each word is tagged with its order in the sentence (first, second, third word, etc.), the number of the word from which it is dependent, and its syntactical role in the sentence (number|dependency|role). In Example 4, 1|3|JCT describes the adverb *yukkuri* 'slowly' as the first word in the sentence, standing in a junctive relation to the third word in the sentence, the main verb. The second word *miruku* 'milk' is the object of the main verb and is thus described as 2|3|OBJ. The verb is the ROOT of the sentence and its dependency is set to 0. The sentence ends with an exclamation mark, which is coded as PUNCT(uation) and is dependent of the main verb.

(4)

*MOT:	yukkuri	miruku	nonde !
%ort:	ゆっくり	ミルク	飲んで !
%mor:	adv|yukkuri=slowly	n|miruku=milk	v:c|nom-IMP:te=drink !
%gra:	1|3|JCT	2|3|OBJ	3|0|ROOT 4|3|PUNCT

This combination of morphological information on the %mor tier and syntactical information on the %gra tier makes possible complex analyses of argument structure in relation to the distribution of word types, inflections, and syntactic roles. Furthermore, the %gra tier can be used for the automatic computation of syntax development measures, such as the Index of Productive Syntax (IPSyn: Scarborough 1990) which

can be already computed by CLAN (Sagae, Lavie, and MacWhinney 2005). The development of a comparable syntax measure for Japanese is still awaited.

2.3 Kana Kanji script in CHILDES

In 1996, CHILDES switched from the limited 256-character ASCII code to Unicode, an international standard defining digital codes for virtually all existing scripts of the world. This rendered possible the use of the Japanese Kana Kanji script in CHAT transcriptions. In practice, a combination of the main utterance line in Hepburn-style augmented by a Kana Kanji version on a separate tier has proved functional (Example 5; based on Example 2), and is currently the prevailing CHILDES format for Japanese.

(5) *MOT: kirinsan ga ohirugohan tabeta ne: .
%ort: キリンさんがお昼ご飯食べたね ~.

With this format style, the transcript becomes more easily readable to the human researcher, and manual checks or coding are facilitated. Because it is the main line in Latin script that forms the basis for automatic computation, the entry format of the Kana Kanji tier (%ort) is relatively unrestricted and is equivalent to common book script. In many cases, an already existing Kana Kanji transcription is imported to a CHAT file as %ort, and a CHAT-format main line is added in Latin script. The drawback is of course the workload of a double transcription.

Alternatively, the use of Kana Kanji on the main line has become possible, too. In this case, the common Kana Kanji script needs to be adjusted to the CHAT format in the following way: Words have to be separated by single-byte spaces, and meta-linguistic characters like "~" for elongation have to be replaced by half-spaced CHAT symbols (e.g., ね: instead of ね ~). Also punctuation is restricted to single-byte characters (" ." instead of "。"; Example 6).

(6) *MOT: キリンさん が お昼ご飯 食べた ね: .

As the switching from two-byte to single-byte characters during text entry is inconvenient, the CLAN text editor provides an automatic correction for spaces and punctuation. The transcriber types continuously in 2-byte mode, and the meta-characters are replaced automatically with their 1-byte counterparts when the file is saved. At present, Kana Kanji on the main line is not yet supported for JMOR, but it is possible to adapt the JMOR lexicon and grammar files to Kana Kanji input.

2.4 Phonetic script

For analysis of phonological development including disorders, the CHILDES system has created a second major tool called Phon, which works in compliance with the CHILDES XML data format. It supports International Phonetic Alphabet (IPA) transcriptions, media linkage, multiple blind transcription, automatic labeling of data such as syllabification, and systematic comparisons between actual and target phonological forms. Phon also provides a number of analysis commands, adapted specifically to the needs of phonological research. Built-in dictionaries of pronounced forms are available for Catalan, Dutch, British, and American English, French, German, Icelandic, Italian, and Spanish, but not yet for Japanese. Currently two Japanese corpora including nine children are available in Phon format. Phon-formatted data can be converted to CHAT format with the help of PhonTalk.

The use of Japanese data in the CHILDES framework is described in a self-teaching manual (Miyata et al. 2004), as well as in an online manual (Miyata 2012b). In addition to the explanations of the JCHAT format and convenient transcription utilities, an overview of the currently available Japanese data is provided. Also the production of video and audio links, coding and grammatical tagging, and the use of this information by CLAN are explained in detail. The use of Phon is explained in a separate Phon Manual.

3 The impact of CHILDES on Japanese language acquisition research

CHILDES combines the advantages of longitudinal and cross-sectional corpora. Longitudinal data provide reliable, rich data that permits the study of long-term changes in individual development. Cross-sectional data from a large number of participants permits the study of differences between acquisitional patterns, social groups, and clinical types. The ready availability of reliable media-linked data that are already tagged for morphology, syntax, conversational analysis, or phonology information, and the existence of tagging and coding schemes for a range of purposes, make analyses possible that far exceed simple word frequency counts.

3.1 Main corpora

The first research to utilize CHILDES for Japanese data was Morikawa (1987), who computerized parts of the monumental Noji Corpus (Noji 1973–1977) for her doctoral dissertation on case particle acquisition. This corpus, consisting of diary notes of Noji's son Sumihare born in the 1950s, had been available in printed form, but was

Table 1: List of Publically Available Japanese CHILDES Corpora (as of February 2013)

Child-Parent Conversations		
Corpus Name	**Child (Age)**	**Creator (Year)***
Hamasaki	Tar (2;2–3;4)	Hamasaki (2004)
Ishii	Jun (0;6–3;8)	Ishii (2004)
MiiPro – ArikaM	ArikaM (3;0–5;1)	Nisisawa and Miyata (2010)
MiiPro – Asato	Asato (3;0–5;0)	Miyata and Nisisawa (2009)
MiiPro – Nanami	Nanami (~~2;211~~–5;0)	Nisisawa and Miyata (2010)
MiiPro – Tomito	Tomito (2;11–5;1)	Miyata and Nisisawa (2009)
Miyata – Aki	Aki (1;5–3;0)	Miyata (2004a)
Miyata – Ryo	Ryo (1;3–3;0)	Miyata (2004b)
Miyata – Tai	Tai (1;5–3;1)	Miyata (2004c)
Noji	Sumihare (0;0–6;11)	Noji, Naka, and Miyata (2004)
Okayama	130 children between 2;0 and 4;11	Okayama, Miyata, Shirai, and Sakazaki (2013)
Special Corpora		
Corpus Name	**Child (Age)**	**Creator (Year)***
Bilingual – Hayashi	Anders (0;12–2;5) Japanese-Danish	Hayashi (2004)
CA – Sakura	18 conversations of groups of 4 students	Miyata, Banno, Konishi, Matsui, Matsumoto, Ōki, Takahashi, and Muraki (2009)
Frogs – Inaba	90 children 3;0–11;0; 48 JL1 adults; 50 JL2 adults	Inaba (2014)
PhonBank – Ota	Hiromi (1;0–2;0); Kenta (1;5–2;6); Takeru (1;4–2;0)	Ota (2008)

not readily used due to its sheer size: the four volumes contain Sumihare's utterances from birth to his seventh birthday, complete with contextual notes. The utterances (without the contextual notes) were published in CHILDES in 2004.

The Japanese database is continuously growing (see Table 1; a more detailed description can be found in the online manual; Miyata 2012b). One of the most frequently used corpora up to now is the Miyata corpus, published between 1995 and 2000. It includes weekly mother-child interactions of three boys (Aki, Ryo, and Tai) between 1;5 and 3;0, who were living in the Nagoya area. The transcripts of Tai are supplied with audio links. As mentioned above, media linkage grants a higher reliability of the transcripts, and most recent corpora are now linked to audio or video files. For example, the Ishii Corpus, published in 2004, is a video-linked corpus of father-child conversations of the boy Jun (0;6–3;8) who was raised in Kyoto. Similarly, the MiiPro Corpus (2009, 2010) includes audio-linked conversational data

of two girls (Arika and Nanami) and two boys (Asato and Tomito) living in the same neighborhood in Tokyo. The data cover the age span between 1;2 and 5;0, although at present, only the data between 3;0 and 5;0 are publicly available. (See Table 1: Child-Parent Conversations.)

Audio-linkage in combination with the Phon format also facilitates phonetic research. For example, the Ota Corpus (2008) includes orthographic as well as phonetic transcriptions of the utterances of the girl Hiromi (1;0-2;0), and the boys Kenta (1;5–2;6) and Takeru (1;4–2;0). (See Table 1: Special Corpora.)

Other recently added corpora are the Sakura Corpus (2009), the Inaba corpus (2014) and the Okayama corpus (2013). The Sakura Corpus consists of 18 twenty-minute conversations of four college students in various gender combinations. The Inaba Corpus is an audio-linked collection of frog story narrations from 90 children between 3;0 and 11;0, 48 Japanese adults, and 50 second language speakers of Japanese (English native) with proficiency level 1 to 5. Also the English versions of the story are included.

The Okayama Corpus includes data from 130 mother-child pairs in the Osaka region between 1969 and 1971. The data cover all utterances between the child and his or her mother during one whole day, and were collected by Okayama's students with the help of handwritten notes and tape recordings.

3.2 Research based on CHILDES data

Around 1995, when after the first CHILDES workshop took place in Japan, a growing number of research presentations and papers on Japanese language acquisition utilizing CHILDES started to appear. In the beginning, researchers relied on the CHAT format and the CLAN programs to collect new corpora to address their research questions. More recently, the availability of large quantities of fully transcribed data has made it possible for researchers to conduct interesting analyses without having to collect new corpora. But this ability to conduct analyses directly from existing corpora does not mean that researchers will stop collecting new data. Instead, we expect that, going forward, researchers will rely on both new corpus collection and the analysis of existing corpora. The research review below summarizes research using data from the CHILDES database, as well as unpublished data in CHAT-format collected by the respective author, which case no corpus name is mentioned.

3.2.1 Research on case and argument structure

A great proportion of research using CHILDES dealt with the acquisition of case and argument structure. Morikawa (1987, 1997) investigated how Sumihare (Noji Corpus) started to use case particles with transitive and intransitive verbs. Nishibu (1998,

2000) investigated the use of the nominative case particle *ga* with Aki from the Miyata Corpus and Yū, a boy aged 1;6-3;0 observed by Nishibu herself. Matsuoka (1998a, 1998b, 2001) studied the acquisition of the case particles *ga* (nominative), *o* (accusative), and especially *ni*, which is used as dative case particle and as locative postposition, on the basis of the speech data of Aki, Sumihare (Noji Corpus), and her own data from a boy called Kan. Similarly, Sugisaki (2011) investigated the difference of case particles and postpositions in the course of acquisition (Sugisaki, 2011) using the data of Aki and Tai from the Miyata Corpus, and Sugisaki (2005, 2008) focused on word order parameter setting including the third child, Tai, from the Miyata Corpus. Tanaka (2011) and Tanaka and Shirai (in press) investigated word order, case particles, and animacy as cues for argument structure in the speech data of Asato and Arika (MiiPro Corpus) and their mothers.

Miyamoto et al. (1999) focused on the high omission rate of the case particles *ga*, (nominative), *o* (accusative), and the topic particle *wa* using the Aki corpus. Also Hirakawa, Oshima, and Itoh (2009) investigated the omission pattern of the nominative case particle *ga* in an unnamed girl aged 2;l–2;11, whose data were transcribed in CHAT format. Guerriero, Oshima-Takane, and Kuriyama (2006) investigated the distribution of null-argument, pronominal argument and lexical argument in English and Japanese data from a discourse-pragmatic angle including pointing and gaze utilizing the coder mode of CLAN format. Also Fujimoto (2008), who analyzed the data of Aki, Ryo, and Tai, took a pragmatic look at the early use of case and topic particles. Kayama (2006) conducted an experimental study (transcribed and coded in CHAT format) with from ten Japanese-speaking children between 2;5 and 4;0 concerning the children's interpretation of null-arguments.

3.2.2 Research on attributive phrases and clauses

Also, relative clauses and attributive phrases were investigated intensively. Ozeki (2005, 2008) and Ozeki and Shirai (2005, 2007a, 2007b, 2010) explored the acquisition of relative clauses comparing the data from Aki, Ryo, Tai, Sumihare (Noji Corpus) and Taro (Kokuritsu Kokugo Kenkyūjo, 1982a, b) to Korean data. Sugisaki (2010) analyzed the input for the same data (Aki, Ryo, and Tai) from a generative point of view, focusing on the different structure of English and Japanese relative clauses. Murasugi, Nakatani, and Fuji (2011) reinvestigated the overuse of the genitive case particle *no* in attributive clauses and the acquisition of attributes, based on the data of Jun (Ishii corpus [see Table 1]) and Sumihare from the CHILDES database, and a third boy called Yuta whose data were collected by the authors.

3.2.3 Research on complex predicates and compound nouns

Miyoshi (1999) explored the timing of the acquisition of complex predicates and compound nouns by Aki from a UG point of view. Also, Murasugi (2011) investigated

frequently produced typical grammatical errors in a parameter framework, on the basis of Sumihare's speech data, especially case marking errors, overgeneration of the complementizer *no*, root infinitive analogues (the early use of past tense *-ta* in volitional contexts) and verb inflection errors involving problems with transitivity and causativity.

3.2.4 Research on verb types and verb inflection

Other studies on grammar acquisition focused on verb types and verb inflection. Nomura and Shirai (1997) investigated the proportion of intransitive and transitive verbs on the basis of CHAT-formatted observational data of one Japanese-speaking boy and his mother. Fuji (2006) and Murasugi, Hashimoto, and Fuji (2007) investigated the acquisition of transitivity, intransitivity, nonaccusativity, and causativity on the basis of Sumihare's speech data. Also, Fukuda and Choi (2010) compared the proportion of transitive verbs, intransitive verbs, and adjectives in the speech Aki, Ryo, Tai, and Jun to the data of four Korean-speaking children and their mothers.

3.2.5 Research on tense and aspect

The acquisition of tense and aspect was a topic in first language as well as in second language acquisition. Shirai (1994, 1995) examined the developmental changes of tense-aspect marking in adult learners of Japanese, while Shirai (1998) and Shirai and Li (2000) compared the acquisition of tense and aspect of three children, namely Aki, Taro (Kokuritsu Kokugo Kenkyūjo, 1982a, b), and the a third child named Yotchan, who had been observed by Patricia Clancy. Shirai and Suzuki (2012) further analyzed the factors determining the acquisition of aspect including verb type and input frequency on the basis of the Tar, Jun, and Ryo Corpora. Shirai and Miyata (2006) focused on the acquisition of the past tense morphology on the basis of Aki, Ryo, and unpublished diary data of a girl called Kī by Masayuki Yokoyama. A follow-up study by Kubo and Suwa (2008) reanalyzed the past tense use of Aki, Ryo, Tai, proposing an additional third step in the development. Kubo (2009) contrasted the development of past tense use of by the three children to the acquisitional process of English-native L2 learners of Japanese.

Other studies dealt with the acquisition causative, negation and other verb inflections. Shirai et al. (2000, 2001) explored the acquisition of the causative combining conversational data from Tai, diary data from the above mentioned diary study of Kī collected by Yokoyama, and cross-sectional data from the Okayama Corpus. Kubota (2011a, b) investigated the acquisition of negation using the Jun Corpus, especially focusing on the effects of dialectal variations in the input (Kansai dialect

-hen vs. Tokyo dialect *-nai*). Klafehn (2004) focused on morphological errors occurring during the acquisition of verb inflection using the Aki, Ryo, and Tai corpora.

3.2.6 Research on early vocabulary: Verbs, nouns, and verbal nouns

A great number of studies focused on the characteristics of early vocabulary, especially verbal nouns, and the proportion of verbs and nouns. For example, Yamashita (1999) examined the composition of the early vocabulary of two Japanese-speaking children and their mothers, comparing the acquisition rate of noun, verbs, and verbal nouns. Oshima, Miyata, and Naka (2000) compared the early vocabulary composition and maternal input of two English-speaking and two Japanese-speaking (Tai and Ryo) children, especially focusing on the verbal and nominal use of English deverbal nouns (e.g., *I drink milk* vs. *Did you get a drink already?*) and Japanese verbal nouns (e.g., *Jon kara denwa ga atta* 'There was a phone call from John' vs. *Denwa ga otita* 'The phone has fallen down'). Miyata, Oshima-Takane, and Nisisawa (2004) analyzed the early vocabulary of four children and their mothers (MiiPro corpus). Miyata (2012c) reinvestigated the vocabulary of two of the four MiiPro children, Nanami and Arika, who had displayed an extremely high, respectively low, noun bias. Ogura (2006) and Ogura et al. (2006) examined the role of the situation for the frequency of verbs and nouns using cross-sectional data of 31 children between 1;0 and 2;0 in book-reading and in toy-play situations. The same data were used in their study of the connection between lexical and grammatical development as measured by MLU (Ogura et al. 1997), and in Ogura, Yamashita, and Tsubota (1997) and Ogura (2006), who they focused especially on the use and function of baby-talk words. Suzuki (2009, 2013) investigated the acquisition of verbal nouns on the basis of the Aki, Ryo, Tai, Tar, Jun, and Sumihare Corpora. She concentrated on the syntactical frame in which early baby talk verbal nouns appeared, highlighting the acquisition process of transitive and intransitive verbal nouns (*nainai [suru]* '[do] putting away' vs. *nenne [suru]* '[do] sleeping').

3.2.7 Research on pragmatic functions

A number of studies focused on the pragmatic aspects functions of language acquisition. Itoh and Oshima-Takane (2004) and Itoh (2008) investigated the acquisitional process of deictic words (so-called *ko-so-a* words) on the basis of the speech data from Aki, Ryo, and Tai. Itoh (2006, 2008) compared these results to the developmental data obtained from autistic and high-functioning autistic children. Guerriero, Oshima-Takane, and Kuriyama (2006) compared focused on the pragmatic use of deictics, in comparison to lexical referents and null-arguments, using CHILDES-

formatted data by of six English- and six Japanese-speaking children and their mothers at two time points, 1;9 and 3;0.

Kubota (2000) examined the developmental changes in the turn-taking pattern of a toddler and his mother. Kubota (2010) concentrated on the responses to clarification requests by Jun and an additional boy comparing them to data from English-speaking children. Hamasaki and Shirai (2000) investigated the use of clarification requests in a two year-old girl named Kokoro. Kubota (2010) concentrated on the responses to clarification requests by Jun and an additional boy, comparing them to data from English-speaking children. Hamasaki (2003) investigated the timing of Y/N answers of three children (Tai, Tar, and the aforementioned girl Kokoro), by analyzing the time information of the audio links.

Kasuya and Uemura (2005, 2011) investigated triadic family interactions between each of the parents and two siblings of nine families at two time points when the younger child was 2;7 and 3;1. Tsuji and Stainthorp (2008) focused on the communication style of the mothers of 10 children between 1;1 and 2;1. Kurumada and Iwasaki (2011) focused on the functions of *ii* 'good' in mother-child conversations of Aki, Ryo, and Tai. Miyata and Nisisawa (2007) examined the emergence of backchanneling behavior of Tai, and Nomura (2007) focused on the pragmatic aspects of postposing using the Tai and the Jun corpora.

The Theory of Mind (ToM) is defined as the ability (1) to attribute mental states, such as beliefs, intents, desires, pretending, and knowledge to oneself and others, and (2) to understand that others may have different beliefs, desires, and intentions from one's own. Matsui, McCagg, and Yamamoto (2005) and Yamamoto, Matsui, and McCagg (2005) investigated the use of *datte* 'but, after all' by Tai in the frame of the child's developing a ToM. As *datte* is used in situations where the speaker disagrees with some other opinion, this shows that children understand that people can have differing opinions. In an additional experiment, they examined the use of *datte* by 16 preschool children in comparison to their results in a ToM false-belief test. Matsui, Yamamoto, and McCagg (2006) further extended their study to Aki and Ryo, and examined the use of mental verbs and final particles expressing certainty (*sitte iru* 'I know', assertive particle *yo*) versus verbs and particles expressing uncertainty (*omou* 'I think', doubt expressing particle *kanaa*), and direct knowledge (*miru* 'I see'; assertive particle *yo*) versus hearsay (*kiku* 'I hear', quotative particle *tte*).

Other research focused on how language typology affects the acquisition of the concept of number. Sarnecka et al. (2007) conducted a cross-linguistic study on the acquisition of the numbers one, two, three by children between 2;6 and 3;6 acquiring Japanese, Russian or English, investigating the influence of the availability of singular/plural inflections. All data were derived from CHILDES; the Japanese data included Aki, Ryo, Jun, and Sumihare, and their mothers, as well as experimental data including 48 Japanese children between 2;9 and 3;6.

3.2.8 Research on narratives and book reading

Pragmatic studies that use CHILDES-formatted data also include research on narration and book reading. Minami (1994, 1995) and Minami and McCabe (1995) explored cross-linguistic and cross-cultural differences in the way caretakers elicit narratives about past events by using an unpublished corpus of 8 Japanese- and 8 English-speaking preschool children and their mothers. Minami (1996a, b, 1997, 1998, 2001, 2002a, b) focused on the development of the children themselves analyzing narratives of past events ("injury") of 10 four-year-old and 10 five-year-old children.

Minami further applied Labovian methodology (Labov 1972, 1997) to L2 and bilingual narrative development. Minami (2006, 2009) examined the past event narratives of 32 second-language learners of Japanese in comparison to adult native speakers of Japanese. Likewise, focusing on 40 English-Japanese bilingual children's narrative development, Minami (2008, 2011) examined narrative contents, narrative marking such as the use of the past tense, and cohesive referencing devices. Extending his research to a bilingual and multicultural context, Minami illustrated that narrative characteristics vary by language and culture and that bilingual children's narratives reflect the culture learned at home often even when they are using another language.

When it comes to the development of children's language skills, oral narratives in particular serve as the platform for transition into literacy. In terms of the orality-literacy continuum, Minami (2000a, 2001) investigated the book reading style of Japanese mothers towards their preschool children based on data of 20 mother-child dyads, comparing the amount of immediate (e.g., labeling) and non-immediate talk (real world connections and explanations) and turn sequences.

3.2.9 Research on phonology

The advent of Sonic Chat in combination with the Phon program also rendered possible phonological studies. Ota (1998, 1999, 2001, 2003) investigated the phonological development from the viewpoint of Optimality Theory. Optimality Theory (Prince and Smolensky 2004) is a general model of how grammars are structured. The theory's central idea is that surface forms of language reflect resolutions of conflicts between competing constraints. Based on a rich database from nine children including Aki, Ryo, Tai, Sumihare from the CHILDES database, Taro (Kokuritsu Kokugo Kenji Kenkyūjo 1982a, b), and the girl Y followed by Ai Okubo, as well as subsequently published data of the three children Hiromi, Kenta, and Takeru, he investigated the prosodic characteristics and phonological errors of the children. In Ota (2006, 2013), he explored the relationship between word truncation rate by the children and the frequency of the corresponding word in maternal speech.

4 Current Trends at Japanese CHILDES: Challenges and Perspectives

4.1 Advantages and possibilities

As we have reviewed in the previous section, the high number of CHILDES related research articles shows the enormous impact of CHILDES on acquisitional research. The slightly outdated childesbib.pdf list (available at http://talkbank.org/usage/) includes 3104 publications up to 2008). In our view, CHILDES is especially successful because it responds to the demands of a) reliability, b) replicability, c) variety and flexibility, and d) commitment to data sharing.

a) Reliability

The use of computational transcription aides enhances the quality of transcripts and helps to eliminate errors like typing mistakes. The easy access to the media allows direct checking and correcting of questionable transcriptions. Furthermore, audio/video linkage allows every user to access the original sound data and judge by herself, thus making the analysis more reliable. And of course, the reliability of research results rises considerably when the speech data used in that research are published afterwards and become freely accessible to everyone. And last but not least, the sheer amount of data minimizes the influence of random transcription errors on the analysis.

b) Replicability

Replicability of results is another important requirement in scientific research. The publication of the data, on which a given research analysis is based, permits an exact replication of the research procedure, better control of various factors influencing the analysis results, and the extension to other data.

c) Variety

First of all, the variety of languages covered by CHILDES allows more and more sophisticated cross-linguistic analysis based on large speech corpora, like for example the article on the influence of language type on number acquisition mentioned above (Sarnecka et al. 2007). Equally important is the variety of linguistic research approaches. Speech data are analyzed from various angles: grammatical, phonological, semantic, gestural, mimetic research, analysis of sign language and bilingual data, conversational analysis, and so forth. All these areas can profit enormously from specifically tailored computer tools. CHILDES responds to these needs by the flexible incorporation of sub-programs like Phon or Grasp, developed by members of the CHILDES community, and the support of standard coding and annotation systems like CA coding or syntax coding schemes. CHILDES is also coping with the difficult task of making all the programs and schemes mutually compatible.

d) Commitment to data sharing

CHILDES is a non-profit research system that is based largely on voluntary contributions from researchers and support from public funds. All corpora and programs are freely available to the public, and personal support is provided at no charge. The free and easy access via the Internet renders possible the use of the system by researchers and students all over the world who would not be able to afford a fee-based usage, and facilitates linguistic research already on the undergraduate level.

The core principle underlying the whole system is the expectation that researchers who have been supported through public and university funds and who have used CHILDES programs and data, should eventually contribute their data, the corresponding media, and newly developed annotation schemes or programs to CHILDES. Funding agencies and professional societies now uniformly agree that data collected with public funding should be made available to the public. To facilitate this process, CHILDES provides a flexible frame for data publication (including a citation format that makes a corpus publication equal to other written papers), and provides technical updates and maintenance to already published corpora. It is remarkable how many of the studies mentioned in the research review above used CLAN with data that have not yet been contributed to CHILDES. Encouraging the publication of these "sleeping" data is a major task for the near future. Inclusion of these data in CHILDES may require increased support for data publication and readily available help in use of CLAN for transcription and linkage to media.

4.2 Challenges

A serious drawback of any speech database is the workload of data entry. In the case of Japanese, the typing process takes much more time, because of the Kana Kanji selection procedure (*moji henkan*). When we include the time for typing of an additional Latin script tier, the number of keystrokes increases fivefold compared to Latin script.[1] Of course, any adding of CHAT symbols or codes and tags takes additional time. Although this process is considerably assisted by an array of entry utilities, the transcription time is unlikely to shrink unless reliable automatic speech recognition becomes more generally available. Nevertheless, the number of Japanese corpora keeps growing, and at present (September 2013) 13 Japanese corpora containing a total of 1,103 files and 2,218,290 words are available.

Another challenge for CHILDES in Japan is the language barrier posed by the use of English as a medium for CHILDES: its homepage, the manuals, and the CLAN interface are only available in English. This poses a barrier to Japanese scholars

1 When typing Example 1, for example, the number of keystrokes is 13 in Latin script, and 37 (42) in Kana Kanji script without or with spaces, respectively. The stroke rate can be much higher according to the specific words used in the sentence.

that should not to be underestimated. In the research review about it was striking that most studies were performed by English-Japanese bilingual researchers either located in an English speaking country, and/or having an international background, and a rather high number of studies deals with English-Japanese L2 or bilingual data. Although a Japanese version of the CLAN Manual was early available and updated several times later on (Oshima-Takane and MacWhinney 1995; Oshima-Takane et al. 1998; Miyata et al. 2004; Nomura 2008, Miyata 2012b), the use of CHILDES is not as popular as might be expected. For a future expansion of the Japanese section of CHILDES it is indispensable to make the access to CHILDES easy enough to be used spontaneously by students at undergraduate and postgraduate level. Intensified publicity efforts, including the use of social media like Wikipedia and Facebook, might be effective as well.

Overall, it can be summarized that CHILDES in Japan has developed in many ways, despite the problems posed by a continuous non-Latin script. Several computer tools facilitating data entry and analysis have been developed especially for Japanese, and new Japanese corpora are steadily being published. Nevertheless, the amount of data available is still far from being sufficient.

By way of summary, we can conclude that CHILDES in Japan has developed in many important ways, despite the problems posed by a continuous non-Latin script. Researchers now have powerful computer tools for facilitating data entry and analysis, and these tools have been customized specifically for Japanese. The major current weaknesses in CHILDES for Japanese are not technical ones involving the programs, but limitations in the size and coverage of the database. Although the database continues to grow, the pace of this growth has not kept up with the demands of modern day corpus linguistic analysis, which requires increasingly larger samples from a great number of children across a longer age range in a greater variety of social situations. Research results analyzing a larger number of Japanese children over a longer period of time is still sparse. Furthermore, research based on the Japanese CHILDES corpora has often combined one or two children from the database with the researcher's own unpublished data. This means that published results are often based in part or entirely on unpublished data. Until the data for such analyses is made publicly available, the scientific basis for Japanese language acquisition research remains insecure. In other words, Japanese acquisition research has not yet entered the age of "big data".

The core problem here is the unwillingness of Japanese child language researchers to make their data publicly available. In our view, the research community needs to consider data publication as the standard procedure, rather than as an optional process. This is particularly true for research supported by national grants. Making the results of these studies available to the research community, under password protection if necessary, would greatly enhance the reliability of research results, while laying the foundation for the advent of a real database age in the study of Japanese language learning.

References

Brown, Roger. 1973. *A first language*. Cambridge, MA: Harvard University Press.

Coulmas, Florian. 1989. *The writing systems of the world*. Malden, MA: Blackwell.

Fuji, Chisato. 2006. Two types of causatives in Japanese and Japanese sign language: A study in syntax and acquisition. Nagoya: Nanzan University master thesis.

Fujimoto, Mari. 2008. L1 acquisition of Japanese particles: A corpus-based study. New York: The City University of New York dissertation.

Fukuda, Shin and Soonja Choi. 2010. The acquisition of transitivity in Japanese and Korean children. In Shoichi Iwasaki, Hajime Hoji, Patricia M. Clancy and Sung-Ock Sohn (eds.), *Japanese Korean Linguistics* 17, 613–624. Stanford, CA: CSLI Publications.

Guerriero, A. M. Sonia, Yuriko Oshima-Takane and Yoko Kuriyama. 2006. The development of referential choice in English and Japanese: a discourse-pragmatic perspective. *Journal of Child Language* 33(4), 823–857.

Hamasaki, Naomi. 2003. The timing-shift of two-year-old's responses to caretaker's yes/no questions. In Yasuhiro Shirai, Hiromi Kobayashi, Susanne Miyata, Kei Nakamura, Tamiko Ogura and Hidetosi Sirai (eds.), *Studies in Language Sciences* (2), 193–206. Tokyo: Kurosio Publishers.

Hamasaki, Naomi and Hidetoshi Shirai. 2000. A contextual analysis of a Japanese two-year-old's clarification requests. In Eve V. Clark (ed.), *The Proceedings of the Thirtieth annual Child Language Research Forum*, 179–186. Stanford, CA: Center Study Language and Information.

Hirakawa, Makiko, Yuriko Oshima-Takane and Keiko Itoh. 2009. Case omission in early child language. In Shunji Inagaki, Makiko Hirakawa, Setsuko Arita, Yasuhiro Hirakawa, Hiromi Morikawa, Mineharu Nakayama, Hidetosi Sirai and Jessika Tsubakita (eds.), *Studies in Language Sciences* (8), 79–91. Tokyo: Kurosio Publishers.

Itoh, Keiko. 2006. Shijishi ko, so, a no shutsugen kara mita kōkinō jiheishōji ni okeru goyōronteki kinō no tokuchō [Characteristics of the pragmatic functions of the *ko/so/a* deictic words with high-functioning autistic children]. *Japanese Journal of Communication Disorders* 23(3), 169–178.

Itoh, Keiko. 2008. *Jiheishōji no goyōronteki nōryoku ni kansuru jisshōteki kenkyū* [Empirical resarch on the pragmatic abilities of autistic children]. Tokyo: Kazama Shobo.

Itoh, Keiko and Yuriko Oshima-Takane. 2004. Shijishi ko, so, a no shutsugen junjo o kitei suru yōin [Determinants of developmental order of Japanese ko/so/a deictic words]. *Japanese Journal of Communication Disorders* 21(1), 1–14.

Kasuya, Hiroko and Kayoko Uemura. 2005. Forms of parental directives in triadic family interactions. In Masahiko Minami, Kobayashi, Harumi, Mineharu Nakayama and Hidetosi Sirai (eds.), *Studies in Language Sciences* (4), 119–135. Tokyo: Kurosio Publishers.

Kasuya, Hiroko and Kayoko Uemura. 2011. Children's self-talk as internal dialogue in the context of triadic parent-sibling interaction. In Setsuko Arita, Yuko Goto Butler, Eric Hauser, Kaoru Horie, Reiko Mazuka, Yasuhiro Shirai and Jessika Tsubakita (eds.), *Studies in Language Sciences* (10), 79–95. Tokyo: Kurosio Publishers.

Kayama, Yuhko. 2006. The L1 acquisition of Japanese null objects in discourse: a preliminary study. In Mineharu Nakayama, Masahiko Minami, Hiromi Morikawa, Kei Nakamura and Hidetosi Sirai (eds.), *Studies in Language Sciences* (5), 55–73. Tokyo: Kurosio Publishers.

Klafehn, Terry. 2004. Japanese verbal inflection: words and rules or words and words? In Masahiko Minami, Harumi Kobayashi, Mineharu Nakayama and Hidetosi Sirai (eds.), *Studies in Language Sciences* (3), 103–120. Tokyo: Kurosio Publishers.

Kokuritsu Kokugo Kenkyūjo. 1982a. *Yōji no kotoba shiryō (3): Issaiji no kotoba no kiroku* [Child Language Data (3): A Record of Language by a One-year-old]. Tokyo: Shūei Shuppan.

Kokuritsu Kokugo Kenkyūjo. 1982b. *Yōji no kotoba shiryō (4): Nisaiji no kotoba no kiroku* [Child Language Data (4): A Record of Language by a Two-year-old]. Tokyo: Shūei Shuppan.

Kubo, Kazumi and Michiko Suwa. 2008. Early use of the past marker *-ta* in L1 speakers of Japanese. In Timothy Jowan Curnow (ed.), *Selected papers from the 2007 Conference of the Australian Linguistic Society*. http://www.als.asn.au/proceedings/als2007/kubosuwa.pdf (accessed 5 November 2012).

Kubo, Kazumi. 2009. The acquisition of the meanings and uses of the past form *-ta* by first and second language learners of Japanese. Perth: University of Western Australia dissertation.

Kubota, Mihoko. 2000. Development of a Japanese two-year-old's conversational skills through turn-taking with her mother. In Yuriko Oshima-Takane, Yasuhiro Shirai and Hidetosi Sirai (eds.), *Studies in Language Sciences* (1), 135–143. Nagoya: Chukyo University.

Kubota, Mihoko. 2010. An English-Japanese comparative study of clarification responses of parent-child dyads. In Makiko Hirakawa, Shunji Inagaki, Setsuko Arita, Yuko Goto Butler, Kaoru Horie, Eric Hauser, Yasuhiro Shirai and Jessika Tsubakita (eds.), *Studies in Language Sciences* (9), 77–92. Tokyo: Kurosio Publishers.

Kubota, Mihoko. 2011a. A case study of the acquisition of the negative expression *X to chigau*. In Kyōko Sato, Hiroko Ikawa, Yoshie Suzuki, Takako Furuya, Akemi Matsuya, Haruko Miyakoda and Yoshiko Morita (eds.), *Kotoba no shinjitsu o mitsumete: Gengo kenkyū no riron to jisshō* [Watching linguistic facts: Theory and practice of linguistic studies], 251–263. Tokyo: Kaitakusha.

Kubota, Mihoko. 2011b. Aru yōji no dōshi keiyōshi hiteikei no shūtoku: Hyōjungo to Kyōtoben no heiyō kara [A child's acquisition of verbal and adjectival negative forms in the standard form and Kansai dialect of Japanese]. *Proceedings of the Eleventh Annual Meeting of the Japanese Cognitive Linguistics Association* 11, 512–518.

Kurumada, Chigusa and Shoichi Iwasaki. 2011. Negotiating desirability: The acquisition of *ii* 'good' in mother-child interactions in Japanese. In Ho-Min Sohn, Haruko Minegishi Cook, William O'Grady, Leon A. Serafim and Sang Yee Cheon (eds.), *Japanese Korean Linguistics* 19, 511–525.

Labov, William. 1972. *Language in the inner city: Studies in the Black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Labov, William. 1997. Some further steps in narrative analysis. *Journal of Narrative and Life History* 7(1–4). 395–415.

Lee, Laura L. 1974. *Developmental Sentence Analysis. A grammatical assessment procedure for speech and language clinicians*. Evanston: Northwestern University Press.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum. http://childes.psy.cmu.edu/ (accessed 5 November 2012).

Masuoka, Takashi and Yukinori Takubo. 1995. *Kiso Nihongo Bunpō 2nd Edn*. [Basic Japanese Grammar]. Tokyo: Kurosio Publishers.

Matsui, Tomoko, Peter McCagg and Taeko Yamamoto. 2005. Little persuaders: Japanese children's use of *datte* and their developing theory of mind. In Andrea E. Tyler, Mari Takada, Yiyoung Kim and Diana Marinova (eds.), *Language In Use: Cognitive and Usage-based Approaches to Language and Language Learning. Selected papers from Georgetown University Roundtable on Languages and Linguistics 2003*, 36–49. Washington, DC: Georgetown University Press.

Matsui, Tomoko, Taeko Yamamoto and Peter McCagg. 2006. On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development* 21, 158–173.

Matsuoka, Kazumi. 1998a. The acquisition of Japanese case particles and the theory of case checking. Storrs, CT: University of Connecticut dissertation.

Matsuoka, Kazumi. 1998b. The proto-postposition stage in the acquisition of the Japanese particle *ni*. Memphis, TN: University of Memphis.

Matsuoka, Kazumi. 2001. The acquisition of the Japanese case particle *ni*. In Mineharu Nakayama (ed.), *Issues in East Asian language acquisition*, 129–145. Tokyo: Kurosio Publishers.

Minami, Masahiko. 1994. English and Japanese: Cross-cultural comparison of parental styles of narrative elicitation. *Issues in Applied Linguistics* 5(2), 383–407.

Minami, Masahiko. 1995. Long conversational turns or frequent turn exchanges: Cross-cultural comparison of parental narrative elicitation. *Journal of Asian Pacific Communication* 6(4), 213–230.

Minami, Masahiko. 1996a. Japanese preschool children's and adults' narrative discourse competence and narrative structure. *Journal of Narrative and Life History* 6, 349–373.

Minami, Masahiko. 1996b. Japanese preschool children's narrative development. *First Language* 16(3), 339–363.

Minami, Masahiko. 1997. Cultural constructions of meaning: Cross-cultural comparisons of mother-child conversations about the past. In Charlotte Mandell and Allyssa McCabe (eds.), *The problem of meaning: Cognitive and behavioral approaches*, 297–345. Amsterdam: North-Holland.

Minami, Masahiko. 1998. Politeness markers and psychological complements: Wrapping-up devices in Japanese oral personal narratives. *Narrative Inquiry* 8(2), 351–371.

Minami, Masahiko. 2000. Book reading styles of Japanese mothers. In Yuriko Oshima-Takane, Yasuhiro Shirai and Hidetosi Sirai (eds.), *Studies in Language Sciences* (1), 145–156. Toyota: Chukyo University.

Minami, Masahiko. 2001. Maternal styles of narrative elicitation and the development of children's narrative skill: A study on parental scaffolding. *Narrative Inquiry* 11 (1), 55–80.

Minami, Masahiko. 2002a. *Culture-specific language styles: The development of oral narrative and literacy*. Clevedon, England: Multilingual Matters.

Minami, Masahiko. 2002b. Narrative as a reflection of culture and consciousness: Developmental aspects. In Ray T. Donahue (ed.), *Exploring Japaneseness: Japanese enactments of culture and consciousness*, 241–261. Westport, CT: Ablex/Greenwood Publications.

Minami, Masahiko. 2006. The development of narrative structure in the acquisition of Japanese as a second language. In Mineharu Nakayama, Masahiko Minami, Hiromi Morikawa, Kei Nakamura and Hidetosi Sirai (eds.), *Studies in Language Sciences* (5), 191–206. Tokyo: Kurosio Publishers.

Minami, Masahiko. 2008. Telling good stories in different languages: Bilingual children's styles of story construction and their linguistic and educational implications. *Narrative Inquiry* 18(1). 83–110.

Minami, Masahiko. 2009. An analysis of narrative communication strategies used by Japanese-as-a-foreign-language learners. *Journal of Japanese Linguistics* 25. 1–15.

Minami, Masahiko. 2011. *Telling stories in two languages: Multiple approaches to understanding English-Japanese bilingual children's narratives*. Charlotte, NC: IAP-Information Age Publishing.

Minami, Masahiko and Allyssa McCabe. 1995. Rice balls and bear hunts: Japanese and North American family narrative patterns. *Journal of Child Language* 22, 423–445.

Miyamoto, Edson T., Kenneth Wexler, Takako Aikawa and Shigeru Miyagawa. 1999. Case-dropping and unaccusatives in Japanese acquisition. In Annabel Greenhill, Heather Littlefield and Cheryl Tano (eds.), *Proceedings of the 23rd Boston University Conference on Language Development*, 443–452. Somerville, MA: Cascadilla Press.

Miyata, Susanne. 1999. Assigning MLU stages for Japanese. *Journal of Educational Systems and Technologies* 9, 81–92. The Audio Visual Center, Chukyo University Nagoya Japan.

Miyata, Susanne. 2003. Wakachi2002 v.2.1. Wakachigaki gaidorain [Word and morpheme separation guideline for Japanese]. http://childes.psy.cmu.edu/morgrams/Wakachi2002 (accessed 5 November 2012).

Miyata, Susanne. 2012a. Nihongo MLU (heikin hatsuwachō) no gaidorain: Jiritsugo MLU oyobi keitaiso MLU no keisanhō [Guideline for Japanese MLU: How to compute MLUw and MLUm]. Kenkō Iryō Kagaku 2, 1–15. http://aska-r.aasa.ac.jp/dspace/bitstream/10638/5113/1/0039-002-201203-1-17.pdf (accessed 5 November 2012).

Miyata, Susanne. 2012b. CHILDES nihongoban: Nihongo-yō CHILDES manyuaru 2012 [Japanese CHILDES: The 2012 CHILDES manual for Japanese]. http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/chapter01.html (accessed 5 November 2012).

Miyata, Susanne. 2012c. Expressive children and the noun bias hypothesis. *Studies in Language Sciences: Journal of the Japanese Society for Language Sciences* 11, 1–23. Tokyo: Kaitakusha.

Miyata, Susanne, Makiko Hirakawa, Ruth Kanagy, Yoko Kuriyama, Brian MacWhinney, Masahiko Minami, Kyōko Murakami, Hiro Yuki Nisisawa, Yuriko Oshima-Takane, Kiyoshi Otomo, Tomohiko Shirahata, Sirai Hidetosi, Junko Shirai, Yasuhiro Shirai, Masatoshi Sugiura and Hiroko Terada. 2006. The development of the CHILDES-based Language Developmental Score for Japanese (DSSJ). In Mineharu Nakayama, Masahiko Minami, Hiromi Morikawa, Kei Nakamura and Hidetosi Sirai (eds.), *Studies in Language Sciences* (5), 75–90. Tokyo: Kurosio Publishers.

Miyata, Susanne, Brian MacWhinney, Kiyoshi Otomo, Hidetosi Sirai, Yuriko Oshima-Takane, Makiko Hirakawa, Yasuhiro Shirai, Masatoshi Sugiura and Keiko Itoh. 2013. Developmental Sentence Scoring for Japanese (DSSJ). *First Language* 33(2), 200–216.

Miyata, Susanne, Makiko Hirakawa, Keiko Itoh, Brian MacWhinney, Yuriko Oshima-Takane, Kiyoshi Otomo, Yasuhiro Shirai, Hidetosi Sirai and Masatoshi Sugiura. 2009. Constructing a new language measure for Japanese: Developmental Sentence Scoring for Japanese. In Susanne Miyata (ed.), *Development of a developmental index of Japanese and its application to speech developmental disorders*. Report of the Grant-in-Aid for Scientific Research (B)(2006-2008) No. 18330141, 15–66. Nagoya: Aichi Shukutoku University.

Miyata, Susanne, Kyōko Muraki, Hiromi Morikawa and Brian MacWhinney (eds.). 2004. *Kyō kara tsukaeru hatsuwa dētabēsu : shoshinsha no tame no CHILDES nyūmon* [A speech database you can use straightaway: CHILDES for beginners]. Tokyo: Hituzi Syobo.

Miyata, Susanne and Norio Naka. 1998. Wakachigaki guideline for Japanese: WAKACHI98 v.1.1. *The Japanese Society for Educational Psychology Forum Report No*. FR-98-003. Tokyo: The Japanese Association of Educational Psychology.

Miyata, Susanne and Norio Naka. 2010. JMOR05.1: The Japanese morphological analysis program based on CLAN. http://childes.psy.cmu.edu/morgrams/Japanese (5 November 2012).

Miyata, Susanne and Hiro Yuki Nisisawa. 2007. The acquisition of Japanese backchanneling behavior: Observing the emergence of aizuchi in a Japanese boy. *Journal of Pragmatics* 39, 1255–1274.

Miyata, Susanne, Hiro Yuki Nisisawa and Kiyoshi Otomo. 2005. Hatsuwa sampuru kara hattatsu hyōka dekiru ka? [Is it possible to evaluate development from speech samples?]. *Gekkan Gengo* 2005 (10), 66–74.

Miyata, Susanne, Yuriko Oshima-Takane and Hiro Yuki Nisisawa. 2004. Noun-bias in early Japanese vocabulary and characteristics in maternal speech. In Masahiko Minami, Harumi Kobyashi, Mineharu Nakayama and Hidetosi Sirai (eds.), *Studies in Language Sciences* (3), 87–101. Tokyo: Kurosio Publishers.

Miyata, Susanne, Kiyoshi Otomo and Hiro Yuki Nisisawa. 2007. Exploring the Developmental Sentence Score for Japanese (DSSJ): A comparison of DSSJ and MLU on the basis of typically developing and delayed children. *The Japanese Journal of Communication Disorders* 24(2), 79–88.

Miyata, Susanne, Kenji Sagae and Brian MacWhinney. 2013. The syntax parser GRASP for CHILDES. *Journal of Health and Medical Science* 3, 45–62.

Miyoshi, Nobuhiro. 1999. Compounds and complex predicates: Japanese evidence for a "global" parameter. In Annabel Greenhill, Heather Littlefield and Cheryl Tano (eds.), *Proceedings of the Annual Boston University Conference on Language Development* 23(2), 453–461.

Morikawa, Hiromi. 1987. Acquisition of case marking and predicate-argument structures in Japanese: A longitudinal study of language acquisition mechanisms. Lawrence, KS: University of Kansas dissertation.

Morikawa, Hiromi. 1997. Acquisition of case marking and argument structure in Japanese. Tokyo: Kurosio Publishers.

Murasugi, Keiko. 2011. Japanese syntax: Implications from language acquisition. In Setsuko Arita, Yuko Goto Butler, Eric Hauser, Kaoru Horie, Reiko Mazuka, Yasuhiro Shirai and Jessika Tsubakita (eds.), *Studies in Language Sciences* (10), 17–39. Tokyo: Kurosio Publishers.

Murasugi, Keiko, Tomomi Nakatani and Chisato Fuji. 2007. VP-shell analysis for the acquisition of Japanese intransitive verbs, transitive verbs, and causatives. *Linguistics* 43, 615–651.

Murasugi, Keiko, Tomomi Nakatani and Chisato Fuji. 2011. A trihedral approach to the overgeneration of *no* in the acquisition of Japanese noun phrases. In Ho-Min Sohn, Haruko Minegishi Cook, William O'Grady, Leon A. Serafim and Sang Yee Cheon (eds.), *Japanese Korean Linguistics* 19, 527–541. Stanford, CA: CSLI Publications.

Murata, Kōji. 1968. *Yōji no gengo hattatsu* [Child language development]. Tokyo: Baifūkan.

Naka, Norio and Susanne Miyata. 1999. CLAN o mochiita nihongo keitaiso kaiseki puroguramu JMOR [The CLAN-based Japanese morphology analysis program JMOR]. *The Japanese Society for Educational Psychology Forum Report No*. FR-99-0001, The Japanese Association of Educational Psychology.

Ninio, Anat, Catherine E. Snow, Barbara A. Pan and Pamela R. Rollins. 1994. Classifying communicative acts in children's interactions. *Journal of Communication Disorders* 27(2), 157–187.

Nishibu, Mayumi. 1998. Nihongo kakujoshi ga no kakutoku ni tsuite: Y-ji no dēta bunseki [About the acquisition of the Japanese case particle *ga*: an analysis of the data from the child Y]. In: Nagoya Daigaku Gengobunka Kenkyūkai (ed.), *Kotoba no Kagaku* 11, 201–224. Nagoya: Nagoya University.

Nishibu, Mayumi. 2000. Nominative case particle in Japanese first language acquisition: analysis of supply rates in two longitudinal data. In Yuriko Oshima-Takane, Yasuhiro Shirai and Hidetosi Sirai (eds.), *Studies in Language Sciences* (1), 111–122. Toyota: Chukyo University.

Noji, Junya. 1973–1977. *Yōji gengo no seikatsu no jittai* I–IV. Hiroshima: Bunka Hyōron Shuppan.

Nomura, Jun. 2007. Japanese postposing as an indicator of emerging discourse pragmatics. In Heather Caunt-Nulton, Samantha Kulatilake and I-hao Woo, *A Supplement to the Proceedings of the 31st Boston University Conference on Language Development*, 1–9. Somerville, MA: Cascadilla Press.

Nomura, Jun. 2008. CHILDES no kiso [Basic CHILDES]. University of Hawai'i at Mānoa. http://childes.psy.cmu.edu/intro/nomura/ (accessed 5 November 2012).

Nomura, Masami and Yasuhiro Shirai. 1997. Overextension of intransitive verbs in the acquisition of Japanese. In Eve Clark (ed.), *Proceedings of the Twenty-Eighth Annual Child Language Research Forum*, 233–242. Stanford, CA: CSLI Publications.

Ogura, Tamiko. 2006. How the use of "non-adult words" varies as a function of context and children's linguistic development. In Mineharu Nakayama, Masahiko Minami, Hiromi Morikawa, Kei Nakamura and Hidetosi Sirai (eds.), *Studies in Language Sciences* (5), 103–120. Tokyo: Kurosio Publishers.

Ogura, Tamiko, Philip S. Dale, Yoshie Yamashita, Toshiki Murase and Aki Mahieu. 2006. The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-play contexts. *Journal of Child Language* 33, 1–29.

Ogura, Tamiko, Norio Naka, Yoshie Yamashita, Toshiki Murase and Aki Mahieu. 1997. Nihongo kakutokuji to bunpō no hattatsu: CLAN puroguramu ni yoru bunseki [Lexical and grammatical development in the Japanese children: Analysis through the CLAN program]. *Bulletin of the Faculty of Human Development (Kobe University)* 4(2), 31–57.

Ogura, Tamiko, Yoshie Yamashita and Mihoko Tsubota. 1997. Baby talk and children's linguistic and cognitive development. *Bulletin of the Faculty of Human Development (Kobe University)* 5(1), 1–14.

Oshima-Takane, Yuriko and Brian MacWhinney (eds.). 1995. *CHILDES manual for Japanese*. Montreal: McGill University.

Oshima-Takane, Yuriko, Brian MacWhinney, Hidetosi Sirai, Susanne Miyata and Norio Naka (eds.). 1998. *CHILDES Manual for Japanese 2nd edn*. The JCHAT Project, Nagoya: Chukyo University.

Oshima-Takane, Yuriko, Susanne Miyata and Norio Naka. 2000. Acquisition of grammatical categories: role of physical objects and input. In Yuriko Oshima-Takane, Yasuhiro Shirai and Hidetosi Sirai (eds.), *Studies in Language Sciences* (1), 97–109. Nagoya: Chukyo University.

Ota, Mitsuhiko. 1998. Minimality Constraints and the prosodic structure of child Japanese. In David J. Silva (ed.), *Japanese Korean Linguistics* 8, 331–344. Stanford, CA: CSLI Publications.

Ota, Mitsuhiko. 1999. Phonological theory and the development of prosodic structure: Evidence from child Japanese. Washington, DC: Georgetown University dissertation.

Ota, Mitsuhiko. 2001. Phonological theory and the development of prosodic structure: Evidence from child Japanese. *Annual Review of Language Acquisition* 1, 65–118.

Ota, Mitsuhiko. 2003. *The development of prosodic structure in early words: Continuity, divergence and change*. Amsterdam: John Benjamin.

Ota, Mitsuhiko. 2006. Input frequency and word truncation in child Japanese: Structural and lexical effects. *Language and Speech* 49, 261–295.

Ota, Mitsuhiko. 2013. Lexical frequency effects on phonological development: The case of word production in Japanese. In Marilyn Vihman and Tamar Keren-Portnoy (eds.), *The emergence of phonology: Whole word approaches, Cross-linguistic evidence*, 415–438. Cambridge, UK: Cambridge University Press.

Ozeki, Hiromi. 2005. Daiichi, daini gengo ni okeru nihongo meishi shūzoku shūtoku katei: ruikeiron-teki apurōchi [The acquisition process of nominal attributes in Japanese L1 and L2: a language-typological approach]. Doctoral thesis, Tokyo: Ochanomizu Women's University.

Ozeki, Hiromi. 2008. *Daiichi, daini gengo ni okeru nihongo meishi shūshokusetsu shūtoku katei* [Acquisition process of the Japanese attributive clause in first and second language acquisition]. Tokyo: Kurosio Publishers.

Ozeki, Hiromi and Yasuhiro Shirai. 2005. Semantic bias in the acquisition of relative clauses in Japanese. In Alejna Brugos, Manuella R. Clark-Cotton and Seungwan Ha (eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development*, Volume 2, 459–470. Somerville, MA: Cascadilla Press.

Ozeki, Hiromi and Yasuhiro Shirai. 2007a. The consequences of variation in the acquisition of relative clauses: An analysis of longitudinal production data from five Japanese children. In Yoshiko Matsumoto, David Y. Oshima, Orrin W. Robinson and Peter Sells (eds.), *Diversity in Language: Perspectives and Implications*, 243–270. Stanford, CA: CSLI Publications.

Ozeki, Hiromi and Yasuhiro Shirai. 2007b. The acquisition of Japanese noun-modifying clauses: A comparison with Korean. In Naomi McGloin and Junko Mori (eds.), *Japanese Korean Linguistics* 15, 263–274. Stanford, CA: CSLI Publications.

Ozeki, Hiromi and Yasuhiro Shirai. 2010. Semantic bias in the acquisition of relative clauses in Japanese. *Journal of Child Language* 37(1), 197–215.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

Sagae, Kenji, Eric Davis, Alon Lavie, Brian MacWhinney and Shuly Wintner. 2010. Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language* 37, 705–729.

Sagae, Kenji, Avon Lavie and Brian MacWhinney. 2005. Automatic measurement of syntactic development in child language. *ACL '05 Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, 197–204. http://childes.psy.cmu.edu/grasp/acl05-IPSyn.pdf (accessed 5 November 2012).

Sarnecka, Barbara. W., Valentina G. Kamenskaya, Yuko Yamana, Tamiko Ogura and Yulia. B. Yudovina. 2007. From grammatical number to exact numbers: Early meanings of "One," "Two," and "Three" in English, Russian, and Japanese. *Cognitive Psychology* 55, 136–168.

Scarborough, Hollis S. 1990. Index of productive syntax. *Applied Psycholinguistics* 11, 1–22.

Shirai, Yasuhiro. 1994. Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition* 16, 133–156.

Shirai, Yasuhiro. 1995. Tense-aspect marking by L2 learners of Japanese. In Dawn MacLaughlin and Susan McEwen (eds.), *Proceedings of the 19th Annual Boston University Conference on Language Development*, Vol. 2, 575–586. Somerville, MA: Cascadilla Press.

Shirai, Yasuhiro. 1998. The emergence of tense-aspect morphology in Japanese: Universal predisposition? *First Language* 18, 281–309.

Shirai, Yasuhiro and Ping Li. 2000. *The acquisition of lexical and grammatical aspect. Studies on Language Acquisition 16.* Berlin: Mouton de Gruyter.

Shirai, Yasuhiro and Susanne Miyata. 2006. Does past tense marking indicate the acquisition of the concept of temporal displacement in children's cognitive development? *First Language* 26, 45–66.

Shirai, Yasuhiro, Susanne Miyata, Norio Naka and Yoshiko Sakazaki. 2000. The acquisition of causative morphology: Why does it correlate with the imperative? In Eve V. Clark (ed.), *Thirtieth Child Language Research Forum Proceedings*, 87–94. Stanford, CA: CSLI Publications.

Shirai, Yasuhiro, Susanne Miyata, Norio Naka and Yoshiko Sakazaki. 2001. The acquisition of causative morphology in Japanese: A prototype account. In Mineharu Nakayama (ed.), *Issues in East Asian language acquisition*, 183–203. Tokyo: Kurosio Publishers.

Shirai, Yasuhiro and Yoko Suzuki. 2012. The acquisition of the Japanese imperfective aspect marker: Universal predisposition or input frequency? In Peter Sells and Bjarke Frellesvig (eds.), *Japanese Korean Linguistics* 20, 281-293. Stanford, CA: CSLI Publications.

Sugisaki, Kōji. 2005. Early acquisition of basic word order: New evidence from Japanese. In: Alejna Brugos, Manuella R. Clark-Cotton and Seungwan Ha, *Proceedings of the Annual Boston University Conference on Language Development* 29, 2, 582–591.

Sugisaki, Kōji. 2008. Early acquisition of basic word order in Japanese. *Language Acquisition* 15, 183–191.

Sugisaki, Kōji. 2010. A note on the structure of relative clauses in child Japanese. In: Hiroki Maezawa and Azusa Yokogoshi (eds.), *The Proceedings of the Sixth Workshop on Altaic Formal Linguistics. MIT Working Papers in Linguistics* 61, 1–9. Cambridge, MA: MIT Press.

Sugisaki, Kōji. 2011. The distinction between case markers and postpositions in early child Japanese: New evidence for children's grammatical conservatism. In: Nick Danis, Kate Mesh, and Hyunsuk Sung (eds.), *Boston University Conference on Language Development Proceedings Supplement* 35, 1–8. http://www.bu.edu/bucld/proceedings/supplement/vol35/35-Sugisaki_BUCLD35.pdf (accessed 6 November 2012).

Suzuki, Yoko. 2009. Otona to kodomo no danwa ni okeru "nenne (suru)" kōbun to "nainai (suru)" kōbun no keishiki to imi: verbal noun (suru) kōbun no shūtoku kenkyū ni mukete [The form and meaning of the *nenne (suru)* construction and the *nainai (suru)* construction in adult-child conversations: Towards a developmental research of the verbal noun (suru) construction]. Tokyo: Tokyo University master thesis.

Suzuki, Yoko. 2013. Intarakushon no naka de tsukawareru onomatope dōshi: oyako danwa no bunseki kara [Onomatopoetic verbs used in interactions: from the analysis of mother-child conversations]. In: Kazuko Shinohara (ed.), *Onomatope kenkyū no shatei: chikazuku oto to imi*. [The range of onomatopoeia research: Approaching sound and meaning], 167–181. Tokyo: Hituzi Syobo.

Tanaka, Nozomi. 2011. L1 Acquisition of Japanese verb argument structure: How do children acquire grammar in the absence of clear evidence? Pittsburgh: University of Pittsburgh master thesis.

Tanaka, Nozomi and Yasuhiro Shirai. In press. L1 acquisition of Japanese transitive verbs: How do children acquire grammar in the absence of clear evidence? In: Seungho Nam, Heejeong Ko and Jongho Jun (eds.), *Japanese Korean Linguistics* 21. Stanford, CA: CSLI Publications.

Tsuji, Hiromi and Rhona Stainthorp. 2008. An investigation of maternal communication styles during conversations with young children: drawing an analysis of the expression of communicative intentions in Japanese dyads. In: Tamiko Ogura, Harumi Kobayashi, Shunji Inagaki, Makiko Hirakawa, Setsuko Arita and Yasushi Terao (eds.), *Studies in Language Sciences* (7), 95–111. Tokyo: Kurosio Publishers.

Ushijima, Yoshitomo and Kaname Moriwaki. 1943. Yōji no bunshō no kōzō [Sentence structure in child language]. Aiiku Kenkyujo (ed.), Yōji no gengo hattatsu [Child language acquisition], 55–79. Tokyo: Meguro Shoten. http://dl.ndl.go.jp/info:ndljp/pid/1069706 (accessed 5 December 2012)

Watamaki, Toru. 1981. Shoki bumpō kakutokuki no hattatsu no sūryōteki tokusei [Quantitative characteristics of the early grammar acquisition stage]. *Proceedings of the 45th Conference of the Japanese Psychological Association*, 506.

Watamaki, Toru. 1993. Nihongo MLU keisan no tame no hatsuwa bunkatsu gaido dai 2 han. [Guideline for the segmentation of Japanese utterances for the computation of MLU, 2nd edition]. Nagoya: Nagoya Hattatsu Shōgai Kenkyūjo.

Yamamoto, Taeko, Tomoko Matsui and Peter McCagg. 2005. Use of the connective "datte" and development of theory of mind. In: Masahiko Minami, Harumi Kobayashi, Mineharu Nakayama and Hidetosi Sirai (eds.), *Studies in Language Sciences* (4), 83–98. Tokyo: Kurosio Publishers.

Yamashita, Yoshie. 1999. The acquisition of nouns and verbs. In: Annabel Greenhill, Heather Littlefield and Cheryl Tano (eds.), *Proceedings of the Annual Boston University Conference on Language Development* 23(2), 741–752.