Author Accepted copy, to appear in Bilingualism: Language and Cognition

Running title: Changes in referential production among bilingual returnee children

Title: Changes in referential production among Japanese-English bilingual returnee children: A five-year longitudinal study

Authors:

MAKI KUBOTA

Department of Language and Culture, UiT The Arctic University of Norway

VICKY CHONDROGIANNI

School of Philosophy, Psychology, and Language Sciences, University of Edinburgh

SATSUKI KUROKAWA

Faculty of Arts and Letters, Tohoku University

STEFANIE WULFF

College of Liberal Arts and Sciences, University of Florida

JASON ROTHMAN

Department of Linguistics and English Language, Lancaster University

Department of Language and Culture, UiT The Arctic University of Norway

Centro de Investigación Nebrija en Cognición, University of Nebrija

Address for Correspondence:

MAKI KUBOTA

Department of Language and Culture, UiT The Arctic University of Norway

Postboks 6050 Langnes

9037 Tromsø

[makikubota5@gmail.com](mailto:makikubota5@gmail.com)

**Abstract**

This study tracked the referential production of 25 Japanese-English returnee children for five years upon their return to Japan from an English-dominant environment (Mean age = 9.72 at the time of return) and compared their referential strategies to 27 Japanese monolinguals and 27 English monolinguals, age-matched to the returnee’s age at time of return. Returnees used more redundant Noun Phrases (NPs) in both languages to maintain references compared to monolingual peers. In English, no changes in NP use were noted over time, but increased exposure in English led to fewer redundant NPs when maintaining references. In their native Japanese (L1), returnees used less NPs for maintaining references and more NPs for reintroducing references, indicating improved reference tracking longitudinally. In sum, returnees’ referential production is more sensitive to L1 re-exposure effects than second language (L2) attrition and crucially, increased L2 exposure minimizes redundant referent production among bilingual returnee children.

**Keywords**

Bilingual returnee children; referential production; narratives; language exposure; Working Memory; longitudinal

1. **Introduction**

Referring expressions encompass linguistic forms like full noun phrases (NPs; e.g., the boy, his frog), pronouns (e.g., he, she, it, they), and proper names (e.g., Mary, John), serving the crucial function of identifying entities in discourse for the listener or reader. In narrative contexts, referring expressions play an important role in (re)introducing and maintaining characters, contributing to effective communication, particularly when the story is produced without mutual knowledge between the speaker and the listener. Several factors are involved in the choice of a referent—from linguistic internal (morpho-syntactic, semantic, and discourse-pragmatic) constraints to individual differences in working memory (WM) and perspective-taking skills (Serratrice & Hervé, 2015). In addition to this, bilinguals need to deal with differentiating the two referential systems which may vary in terms of the types of linguistic forms used for different contexts, as well as their degree of preference and optionality. For instance, Spanish has two pronominal forms: a null pronoun and an overt pronoun to express reference while English has only one option: an overt pronoun. Such language-specific differences allow for examining cross-linguistic influence (CLI) from one language to the other, which has been widely investigated in bilingual children and adults with various language combinations (Sorace, 2011). Moreover, of potential consequence is the fact that most bilinguals are not exposed to the two languages equally, nor do they have balanced proficiency: dual language experience is fluid and dynamic, especially throughout developmental trajectories in childhood (De Houwer, 2021; Paradis, 2023).

The current study examines a special group of bilingual children, namely Japanese-English returnees: children born in Japan who moved to an English-speaking society at a young age, remained there for a significant period (3.77 years on average) and have now returned to Japan. We examined their L1 (Japanese) and L2 (English) production of referential expressions in narratives longitudinally over the course of five years from the point of return to the homeland (Japan) to examine how changes in language exposure influences the development of referential production in both languages. In addition, given that previous work on returnees found that children’s bilingual experience influences their development in WM (Kubota et al., 2020) and WM in turn predicts referential production (Hendriks et al., 2014), we also investigated the effects of WM on referential production in returnees.

Furthermore, in order to determine what the comparative baseline performance looks like, we also examined the referential production of monolingual Japanese and English children, age-matched to the returnee’s age at time of return. This study, to our knowledge, is one of the first to track the referential strategies of bilingual children over an extended period of time—specifically in a unique context in which linguistic/environmental transitions take place during the course of their childhood development. The pattern of language exposure and usage diverges from the typical heritage bilingualism scenario, where an increase in the (former) heritage language (HL) is typically accompanied by a simultaneous decline in L2 exposure and use. Such context allows us to examine how referential expressions are affected by significant changes in language exposure and/or domain general abilities both in terms of re-exposure effects to the L1 and maintenance/attrition of the L2.

* 1. Referential expressions in Japanese and English

When referencing a character in a narrative discourse, it is crucial for the speaker to select a linguistic form that aligns with the cognitive accessibility of the intended referent while taking into account the knowledge of the listener (Ariel, 1996). This level of accessibility varies based on whether the referent is being introduced for the first time or revisited, as well as the recency of their previous mention. Both in Japanese and English, the initial introduction of referents often involves the use of complete Determiner/Noun phrases (henceforth NPs), signifying new information within the discourse. Subsequent references to these characters typically are used with null pronouns in Japanese and overt pronouns in English, indicating that the referent has become shared knowledge between the interlocutors. However, when the referent that was the topic in the previous discourse is replaced by another referent, there is a preference for utilizing a full NP instead of null or overt pronouns in Japanese and English respectively, a referential strategy aimed at avoiding potential ambiguity.

Although Japanese does possess third-person overt pronouns like '*he*' (kare), '*she*' (kanojo), or '*they*' (karera), their use is extremely rare, especially in child discourse (Mishina-Mori et al., 2018). Indeed, Clancy (1980) found that L1 Japanese children (n = 60) did not mention a single overt pronoun in their narratives. Thus, the choice between full NP vs. pronominal forms (including overt and null forms) in Introduction, Maintenance, and Reintroduction contexts is regarded as a universal strategy that applies to both Japanese and English (Clancy, 1980, 1992).

In addition to referential form choices, morphological markers (i.e., particles) are used to signify topic prominence in Japanese—introduction of a new referent is marked by '*ga*' (e.g., gorilla-*ga*) while null pronouns (unless there is potential ambiguity, triggering the use of '*wa*') are used for maintaining topic continuity. To signify known information, the particle '*wa*' is usually used, but in a Reintroduction context where a known referent re-emerges as a topic in the discourse, NPs followed by '*ga*' or '*wa*' are both appropriate particles to use depending on the context. English, on the other hand, employs definite (e.g., the frog, their dad etc.) and indefinite (e.g., a frog, one day etc.) articles to mark new and known information (except in cases when definiteness can be established through world-knowledge; e.g., Mary peeled *the skin* of a banana, see e.g., Chondrogianni et al., 2015 for details examining this phenomenon in monolingual and bilingual children).

* 1. The role of cross-linguistic influence, exposure, and Working Memory

The ability to comprehend and produce appropriate referents in the discourse in bilingual children is modulated by an interplay of several factors—including internal structure of the language (triggering CLI), experiential variables such as language exposure and use, and domain-general abilities such as WM. Individual differences in WM have been shown to predict appropriate use of references in both monolingual and bilingual children (Serratrice & De Cat, 2020). Understanding others' perspectives and tracking discourse elements are tasks that require cognitive effort. When the cognitive capacity is overloaded, speakers may use referring expressions that become challenging for listeners to recover (Vogels et al., 2015) or be over informative and produce full NPs when the use of a reduced form would be more appropriate (Torregrossa et al., 2021).

According to the computational model developed by Hendriks, (2016), WM plays a crucial role in constructing one's discourse model since reference tracking involves not only storing goal relevant information but also constantly updating information based on each context. Indeed, previous studies have found a correlation between WM and choice of referents in monolingual (Hendriks et al., 2014; Kuijper et al., 2015) and bilingual children (Serratrice & De Cat, 2020; Torregrossa et al., 2021). For instance, a study by Whitely and Colozzo (2013) examined referential production in English monolingual children from kindergarten to second grade and tested their updating and short-term memory abilities. Results showed that the ability to update information in one’s WM correlated with appropriate reference production beyond the effects of age, especially in the Maintenance context where the topic of the discourse is continued across utterances. Furthermore, Torregrossa et al.’s (2021) study on bilingual children from age seven to thirteen demonstrated that better updating skills (measured via an N-back task) modulated the production of less over-specified NPs in their narratives, while Serratrice & De Cat (2020) found that verbal WM predicted the use of an appropriate NPs for both bilingual and monolingual children.

Overspecification behavior such as (i) using NPs in contexts where pronominal forms are a more ecologically valid choice and/or (ii) using overt pronouns when null pronouns would be contextually more appropriate, has been documented in adult bilinguals (Belletti et al., 2007; Lozano, 2006; Ryan, 2015; Sorace & Filiaci, 2006) and children alike (Paradis & Navarro, 2003; Serratrice, 2007; Serratrice et al., 2004, 2012; Sorace et al., 2009; Torregrossa & Bongartz, 2018). In terms of (ii), some authors have attributed this redundant use of overt pronouns to CLI (e.g., Belletti et al., 2007; Müller & Hulk, 2001; Serratrice et al., 2004), arguing for influence from a non-pro drop language (e.g., English) to pro-drop languages (e.g., Greek, Italian, Spanish). However, work by Sorace et al. (2009), which tested acceptability of Italian and English pronominal subjects among younger and older English-Italian and Spanish-Italian bilingual children, found that both bilingual groups—regardless of the language combination—over-extended the interpretation of overt subjects (although to different degrees, thus, not precluding any role for CLI or the differences in distribution of overt subject pronouns between Spanish and Italian as noted in Filiaci et al., 2014). This finding, along with supporting studies (Lozano, 2006; Serratrice et al., 2012), suggest that CLI cannot be the sole explanation to such overuse/acceptance of overt pronouns observed in bilinguals, and thus other factors such as language exposure or dominance, as well as general processing costs are hypothesized to play a role (see Sorace, 2011, 2016 for detailed discussions). Indeed, Sorace et al. (2009) found that young English-Italian bilingual children in the UK overaccepted redundant subject pronouns more than those living in Italy, underlining the importance of exposure (albeit measured via language of the community) in explaining referential strategies among bilingual children.

Similarly, studies that report overspecification behavior like (i) in German-Italian (Torregrossa & Bongartz, 2018) and English-Italian (Serratrice, 2007) bilingual children suggest that such pattern cannot possibly be attributed to CLI, since all the language pairs investigated in these studies possess full NPs. In fact, Torregrossa et al. (2021) demonstrates that the production of redundant NPs in Greek among bilingual children (when clitics or null pronouns are sufficient) is predicted by the degree of dominance of experience in the other language (Albanian, English or German). Children who were less dominant in the target language, Greek, produced more redundant NPs in reference Maintenance context in Greek narratives.

In sum, previous studies suggest that several linguistic-internal and -external, as well as cognitive factors are at play in accounting for referential strategies in bilingual children. The current study adds to this body of work by examining referential production in returnee children longitudinally over the course of five years, specifically examining the role that language exposure plays in terms of the changes/development in referential production of both L1 and L2 in returnee children. Although studies on this population are rather limited, the returnee context presents a naturally occurring laboratory that is relatively untapped despite its potential to contribute important data for determining the relative impact of particular variables. Returnees experience a rather abrupt and dramatic shift in exposure to their languages, often at an age where development is still taking place in monolingual children. Furthermore, the shifts in exposure and use move in the reverse direction of the typical heritage bilingual context: L1 Japanese input and use increase significantly later in childhood, while L2 English input and use decline just as sharply and simultaneously.

There is some evidence which shows that linguistic structures involving integration of discourse/contextual information (as in the case of referential expression) is vulnerable to linguistic changes in returnees (Antonova-Unlu et al., 2021; Flores, 2010, 2012; Kaya-Soykan et al., 2023). Flores (2012) examined the attrition of two linguistic domains—verb placement and object expression—in German via an oral production task among Portuguese-German bilinguals who were born in Germany but moved back to Portugal sometime during childhood to adolescence. The findings revealed that object expression, a linguistic phenomenon constrained by discourse information, was more vulnerable to effects of reduced input than verb placement, a structure considered to be purely syntactic, but only for those who left Germany after puberty. Kaya-Soykan et al. (2023) on the other hand focused on the re-exposure effects to the L1 among Turkish-German bilinguals who returned to Turkey. They found that even after many years of return to Turkey, the returnees showed different perception and production of evidentiality markers in Turkish from their monolingual counterparts, suggesting that syntax-discourse interface structures may be more resistant to re-exposure effects to the L1.

1.4 Current study

The current study investigates the changes in L1 Japanese and L2 English referential production among Japanese-English returnee children over the course of five years (across three time points) since their return to Japan from an English-dominant environment. We first compared their L1 and L2 performance at the point of return to their age-matched monolingual peers to gauge what a comparative baseline performance looks like and, thus, how they may differ (or not) from monolinguals. In line with the previous literature, we expected bilingual returnee children’s referential production to diverge from their monolingual peers in both languages, specifically in the Maintenance context where overspecification (producing more NPs) is often observed in bilingual children. We may also find underspecification in the Reintroduction context in both languages of bilingual returnees, in which they produce more ambiguous pronouns instead of full NPs. Such behavior may be attributed to reduced processing speed as a function of being less exposed to each of their languages as compared to their monolingual peers (see Torregrossa & Bongartz, 2023 for further discussions on underspecification behavior in bilinguals). We then examined how their L1 and L2 referential production changes over time and ran additional correlational analyses with relative language exposure and WM performance to probe for individual differences. We predicted re-exposure effects in their L1 and attrition in their L2 over time especially in the Maintenance and Reintroduction contexts. That is, children should use *less* NPs in the Maintenance context and *more* NPs in Reintroduction context for L1 Japanese (indicating target-like referential strategies), and vice-versa in English over time, as these two contexts have been shown to be vulnerable in child bilingual literature and display both underspecfication and overspecification behavior respectively. In terms of the correlational analyses, we expect both language exposure and WM to modulate referential strategies in returnee children both at single time points and across time. There should be a significant correlation between language exposure, WM, and NP use—the more the child is exposed to English/Japanese and the better their WM is, the less they use NPs in Maintenance and the more they use NPs in the Reintroduction context.

**2. Methods**

2.1 Participants

The participants in the current study included 25 Japanese returnee children, tested at three time points: (i) Time 1: a few weeks after their return to Japan (Mean age = 9.72, SD = 1.47, Range = 7.65 – 12.99; 15 Female), (ii) Time 2: around a year after the first test session (Mean age = 10.71, SD = 1.51, Range = 8.68– 14.01), (iii) Time 3: around five years after the first test session (Mean age = 15.16, SD = 1.47, Range = 13.14 – 18.43), as well as 27 Japanese monolinguals (Mean age = 9.75, SD = 1.51, Range = 7.09 – 12.48; 15 Female) and 27 English monolinguals (Mean age = 9.93, SD = 1.34, Range = 6.93 – 12.76; 10 Female) age-matched to the returnee children’s age at first session.

Fourteen returnees lived in a country where English is the societal language (USA, UK, Ireland, Canada) while the other eleven returnees lived in a country where English is not the official, societal language (France, Netherlands, Germany, Israel, Vietnam, China, Malaysia). However, all of the latter group of participants went to an English as a medium of instruction schools and their parents reported that they could not communicate in the societal/third language. Their mean age of onset (i.e., age that they moved abroad) was 4.90 years old (SD = 2.59, Range = 1.00 – 9.73) and their average length of residence abroad was 3.77 years (SD = 2.01, Range = 2.00 – 9.74). All participants had Japanese parents and thus had minimal exposure to English before moving abroad. Upon their return to Japan, they all attended Japanese schools (instead of international or bilingual schools). At least one of their parents had university-level education and they all came from families with high Socio-Economic Status (i.e., workers in large revenue companies).

2.2 Instruments

2.2.1 Language Background Questionnaire

We used the Bilingual Language Experience Calculator (BiLEC) (Unsworth, 2016) to quantify the returnees’ language exposure to L1 Japanese and L2 English at first (Time 1), second (Time 2), third sessions (Time 3). In the first session, we asked them to reflect on their language exposure when they lived abroad, while in the second and third sessions, we asked about their language exposure at the time of testing. The questionnaire was administered via an interview-format with the parents and the child, and they indicated where and with whom the child spent time for how long, and which languages they used when interacting with others. Based on this information, proportion of language exposure to English (vs. Japanese) were calculated via the algorithm provided by BiLEC. Table 1 shows the quantified language exposure to English at first, second, third sessions.

<Insert Table 1 about here>

2.2.2 Working Memory task (N-back)

The N-back task, adapted from Chevalier's (2018) version, assessed the ability to update information in WM and monitor task sets while constantly updating new information.

In this task, children viewed a series of pictures displayed one at a time and were required to press a space key when the current picture matched the one presented *n* trials back. Children were instructed to press the space bar if they recognized a picture that matched the one shown one trial back (1-back), two trials back (2-back), or three trials back (3-back). The participants completed three difficulty levels, each comprising 32 trials. Four different images (smiley face, cat, house, airplane) were used in each level, presented individually for 1500 milliseconds with a preceding 500-millisecond fixation cross. Each picture appeared eight times per level in random order. The sequence of difficulty levels was predetermined to introduce the children to the easiest level (1-back) first and conclude with the most challenging level (3-back). Within each level, there were eight target pictures (matched) and 24 non-target pictures (unmatched). Participants had 1500 milliseconds to respond before the target changed. A green tick appeared as positive feedback when participants correctly identified a matched picture, whereas an incorrect response was indicated by a red cross when participants pressed the space bar for an unmatched picture. Correct scores were assigned when participants pressed the space bar for matched pictures (hit trial) or refrained from responding to unmatched pictures (correct rejection). Incorrect responses occurred when participants failed to press the space bar for matched pictures (miss trial) or responded to unmatched pictures (false alarm). After each block, the screen displayed the total percentages of correct and incorrect responses. Each block was preceded by a practice session, and there were brief breaks between blocks. Accuracy was scored as 1 for correct responses and 0 for incorrect ones. The accuracy was averaged across levels and used for subsequent analyses (Arizmendi et al., 2018; Gangopadhyay et al., 2016; Hansen et al., 2016).

2.2.3 Narrative task

Two wordless picture books *Frog on his own* (Mayer, 1973) and *Frog, where are you*? (Mayer, 1969) were used to elicit English and Japanese narratives respectively. Both stories follow a similar story structure (i.e., the frog goes on an adventure and encounters various characters and events) and *Frog on his own* comprises 30 pages of illustrations featuring twelve characters, whereas *Frog, where are you*? consists of 29 illustrations with seven characters.

2.2.3.1 Coding scheme

A Japanese-English bilingual research assistant transcribed the Japanese and English narratives at first, second, and third sessions according to the CHAT system (MacWhinney 2000) and the entire transcription and coding were double-checked by another researcher who is also a Japanese-English bilingual. Any discrepancies were discussed and resolved among the two transcribers/coders. Both Japanese and English transcriptions were first segmented into communication units (C-units), defined as an independent clause plus its modifier (Loban, 1976). Then all utterances in Japanese were coded for referential context (First Mention, Maintenance, Reintroduction), referential form (null pronoun, NP), and particles (e.g., null, “NP+*ga*”, “NP+*wa*”). For English, we also coded for referential context (First Mention, Maintenance, Reintroduction) and referential form (overt pronoun, NP), as well as nominal markers (indefinite nominals, definite nominals, and pronominals). In terms of the nominal markers, we coded them according to Colozzo & Whitely (2015, p.144): (i) indefinite nominals included nouns preceded by indefinite articles (e.g., a dog) or numerals (e.g., one boy); (ii) definite nominals included nouns preceded by definite and demonstrative determiners (e.g., the dog, this frog), possessive pronouns (e.g., the boy’s frog, their dog), and proper names (e.g., Mark); (iii) pronominals included personal (e.g., she, he, they, it) and possessive pronouns (e.g., hers, his, theirs) as well as ellipses of the NP within coordinated clause structures (e.g., he woke up and ø left the house). The referential context for both languages was coded following Berman & Slobin (1994, p.660-664): (i) First Mention: the speaker mentions a story character for the first time in the entire discourse (ii) Maintenance: the speaker mentions continuous reference to the same character within one utterance, in successive utterances, or across multiple utterances that do not progress the event within a story (iii) Reintroduction: the speaker refers back to a character that was previously mentioned but was not the topic of the previous discourse. If it expressed a subject argument whose immediate antecedent was in object position, this was coded as Reintroduction. Cases where the speaker changed from referring to two or more characters (e.g., cat and dog, them) to only a subset (e.g., cat or dog) and vice-versa were also coded as Reintroduction. The example of the coding scheme is provided in (1).

|  |  |  |
| --- | --- | --- |
| **Sentence** | **Context** | **Form** |
| ***A boy and two frogs******and a dog*** walked to the city park. | Introduction | NP |
| ***The frog*** jumped out of the bucket. | Reintroduction | NP |
| And ***Ø*** landed on the path. | Maintenance | Null pronoun |
| ***He*** waved goodbye to the boy. | Maintenance | Overt pronoun |
| ***The frog*** spied a fly. | Reintroduction | NP |
| **He** caught it on his tongue. | Maintenance | Overt pronoun |

(1)

2.3 Procedure

The returnee participants were recruited via information sessions provided by the Japan Overseas Education Services (JOES) organization for returnee families. The first and second test sessions were conducted in-person with a Japanese-English bilingual researcher at the participants’ home or in JOES classrooms. The third test session was conducted online via Zoom with the same researcher. We counterbalanced the order of the narrative tasks (Japanese vs. English) and the WM task (N-back) always took place in between the two narrative tasks. The N-back task for the first and second test session was implemented via E-prime, while in the third test session, we constructed the same task using the online experiment builder platform, Gorilla. The instructions were given in the respective languages and the participants were told that their narratives will be recorded, and their peers will be listening to their recordings who have never seen the wordless picture books. In the first and second test sessions for the returnees and Japanese and English monolinguals (which took place in-person), the researcher sat across from the participants so they could not see the picture book. As for the third test session for the returnees which took place online, the researcher sent a link to the participant via Zoom which displayed the pictures. Identical measures in terms of administering the task were taken for the first, second, and third test sessions.

The Japanese monolingual data were collected in Japan and the English monolingual data in the UK. All testing sessions in both countries were conducted in-person at the participants’ home or at a community center with a Japanese-English bilingual researcher. The procedure for eliciting the narrative tasks were identical to that of the returnees.

1. **Results**

3.1. Descriptive analysis of particles in Japanese and articles in English

As an initial step, we will descriptively analyze the use of particles and nominal markers in Japanese and English respectively to examine what kind of morphological markers are used with the referent in each context. In Japanese, we limited our analysis to the two main particles: *ga*, *wa*, as well as null pronouns (for this descriptive analysis only). As illustrated in Figure 1, we see the expected pattern: (i) NP+*ga* is used most frequently in the First Mention context (ii) null pronouns are used most frequently in the Maintenance context followed by NP+*wa* (iii) NP+*wa* is used more frequently than NP+*ga* in the Reintroductioncontext*,* but this difference is especially prominent at Time 3.

<Insert Figure 1 about here>

In English, as illustrated in Figure 2, returnee children (i) use both definite and indefinite markers to a similar degree in the First Mention context (ii) use pronominal forms most frequently followed by definite markers in the Maintenance context (iii) use definite markers most frequently followed by pronominal forms in the Reintroduction context.

<Insert Figure 2 about here>

3.2. Monolinguals and returnees at first test session

We now describe the returnees’ referential productions in L1 Japanese and L2 English from the first session, comparing it to that of age-matched monolingual peers in order to elucidate the returnees’ individual and comparative baseline performances. Figure 3 illustrates a box plot of the proportions on NPs produced within each context (First Mention, Maintenance, Reintroduction) for each language/group (Japanese returnee, Japanese monolinguals, English returnee, English monolinguals). We ran a generalized linear mixed effects model with (definite and indefinite) NP use (0,1) as a binary dependent variable and Group (Japanese returnee, Japanese monolinguals, English returnee, English monolinguals) and Context (First Mention, Maintenance, Reintroduction) and the interaction between Group and Context as fixed effects as well as Participant and Item as random intercepts and Context as a by-Participant slope. Please see Table S1 in the supplementary materials for the model output.

<Insert Figure 3 about here>

Post-hoc comparisons using Tukey corrections (see Table 2 for all comparisons) revealed that for the First Mention context, English monolinguals produced less NPs than returnees in English (*E* = -1.28, *p* < .001), returnees in Japanese (*E* = -1.11, *p* = .02), and Japanese monolinguals (*E* = -1.12, *p* = .01). As for the Maintenance context, returnees in English produced more NPs than English monolinguals (*E* = -0.93, *p* < .001) and Japanese monolinguals (*E* = 0.91, *p* < .001) and returnees in Japanese produced more NPs than Japanese monolinguals (*E* = -0.62, *p* = .01) and English monolinguals (*E* = -0.63, *p* = .01). Finally, regarding the Reintroduction context, returnees in English produced more NPs than in Japanese (*E* = 0.79, *p* < .001) as well as English monolinguals (*E* = -0.62, *p* < .001) and Japanese monolinguals (*E* = 0.9, *p* < .001).

<Insert Table 2 about here>

3.3. Longitudinal data from returnees

We will now examine the longitudinal data of returnees from first (Time 1), second (Time 2), and third (Time 3) sessions in English (L2) and Japanese (L1). Figure 4 shows a box plot of the proportions on NPs produced within each context (First Mention, Maintenance, Reintroduction) for each session (Time 1, Time 2, Time 3) in each language (English and Japanese). We ran a generalized linear mixed effects model with NP use (0,1) as a binary dependent variable and Time, Context, Language, and a two-way interaction between Time and Context, Time and Language, and Language and Context, as well as a three-way interaction between Time, Context, Language as fixed effects as well as Participant and Item as random intercepts and Context and Time as a by-Participant slope. Please see Table S2 in the supplementary materials for the model output.

<Insert Figure 4 about here>

Post-hoc comparisons using Tukey correction (see Table 3 for full comparisons) revealed no changes in the proportion of NP use for English across all contexts (*p*’s > .06). As for Japanese, we see no changes in the proportion of NPs for First Mention, but there is a *decrease* in the proportion of NPs from Time 2 to Time 3 (*E* = 0.58, *p* = .01) in the Maintenance context and an *increase* in the proportion of NPs from Time 1 to Time 3 (*E* = -0.44, *p* < .001) and Time 2 to Time 3 (*E* = -0.35, *p* = .02) in the Reintroduction context.

<Insert Table 3 about here>

3.3.1 Correlation between relative language exposure, Working Memory, and proportions of NP use in each context at each time points

In order to examine whether the proportions of NP use in each context correlates with relative language exposure and WM measures (see Table S3 in the supplementary materials for descriptive statistics of the N-back performance) at each time point, we ran a Pearson bivariate correlation as indicated in Table 4. That is, we ran a correlation between (i) relative exposure (higher numbers indicate more exposure to English and less exposure to Japanese) and accuracy on the N-back task (WM) at Time 1 to the proportions of NP use at Time 1 for each context (First Mention, Maintenance, Reintroduction) and language (English, Japanese), (ii) relative exposure and WM at Time 2 to the proportions of NP use at Time 2 for each context and language, (iii) relative exposure and WM at Time 3 to the proportions of NP use at Time 3 for each context and language. Positive Pearson’s *r* values indicate that there is a positive relationship between relative exposure, WM, and NP use in each context—that is, more English exposure (i.e., less Japanese exposure) and better WM is associated with *more* NP use in each context. In contrast, negative Pearson’s *r* values suggest a negative relationship; more English exposure (i.e., less Japanese exposure) and better WM is associated with *less* NP use in each context. We only found a significant correlation between relative language exposure at Time 2 and proportions of NP use at Time 2 in the Maintenance context (*r* = .48, *p* = 0.02) for English as shown in Table 4. This suggests that returnee children who received *less* exposure to English after a year of return to Japan displayed the tendency to be redundant (i.e., higher proportions of NP use) when continuing to refer to a referent that had already been introduced as the topic in previous discourse in English.

<Insert Table 4 about here>

3.3.2 Correlation between the *differences* in relative language exposure, Working Memory, and proportions of NPs across time points in each context and in each language.

We now examine whether there is a correlation between the *changes*/*differences* in relative language exposure, WM, and NP use from Time 1 to Time 2, Time 2 to Time 3, and Time 1 to Time 3 in each context (First Mention, Maintenance, Reintroduction) and language (English, Japanese). In order to compute the difference scores in NP use, we subtracted the proportions of NPs of Time 1 from Time 2 (Interval 1; Time 2 – Time 1), Time 2 from Time 3, (Interval 2; Time 3 – Time 2) and Time 1 from Time 3 (Interval 3; Time 3 – Time 1) in each context and language, so that *positive* values indicate *more use* of NPs over time. As for the difference scores of relative language exposure, we subtracted the (i) relative proportion of exposure of Time 1 from Time 2 (Interval 1; Time 2 – Time 1), Time 2 from Time 3 (Interval 2; Time 3 – Time 2), and Time 1 from Time 3 (Interval 3; Time 3 – Time 1), in which *negative* values indicate *greater loss* of English exposure (i.e., greater gains in Japanese exposure) over time (ii) WM performance of Time 1 from Time 2 (Interval 1; Time 2 – Time 1), Time 2 from Time 3, (Interval 2; Time 3 – Time 2) and Time 1 from Time 3 (Interval 3; Time 3 – Time 1) in which *positive* values indicate *improved* WM performance over time. We then ran a bivariate correlation between (i) difference scores in exposure and WM at Interval 1 to the differences in proportions of NPs at Interval 1 for each context (First Mention, Maintenance, Reintroduction) and language (English, Japanese), (ii) difference scores in exposure and WM at Interval 2 to the differences in proportions of NPs at Interval 2 for each context and language and (iii) difference scores in exposure and WM at Interval 3 to the differences in proportions of NPs at Interval 3 for each context and language, as indicated in Table 5. There was a significant positive correlation between the difference score in exposure at Interval 3 and the difference in proportions of NPs at Interval 3 for the Maintenance context in English (*r* = .40, *p* = 0.04). That is, children who experienced *more loss* in English exposure since their return to Japan used *more redundant NPs* in the English maintenance context over the five-year period since returning to Japan.

<Insert Table 5 about here>

**4. Discussion**

The current study tracked L1 and L2 referential production among Japanese-English returnee children over the course of five years since their return to Japan from an English-dominant environment and correlated their performance to language exposure and WM measures to examine what factors may predict their referential production in L1 and L2 at each time point and longitudinally. We also compared their referential production in L1 Japanese and L2 English to their age-matched monolingual peers to establish their comparative baseline performance.

Firstly, we will discuss the findings of the data from returnees in Japanese and English (at first session) and monolingual Japanese and English children. Interestingly, English monolinguals produced less NPs than Japanese monolinguals in Japanese as well as returnees in both English and Japanese in the First Mention context only, the one in which we might have expected the least amount of difference across groups, irrespective of the language of testing, given discourse constraints on (anaphoric) identification. While the differences noted is indeed statistically significant, this does not mean that the groups are so distinct in the relevant sense. After all, the English monolinguals produced full NPs 94% of the time. Furthermore, the difference does not represent a systematic distinction across the groups when you consider the ranges in each: (i) English monolinguals (78% - 100%) (ii) returnees in English (95% - 100%); (iii) returnees in Japanese (82% - 100%); (iv) Japanese monolinguals (83% - 100%). Of the 27 English monolinguals, all but one produced at least 87% suppliance of NPs. In other words, 26 of the English monolinguals fall within the range of the next most conservative group (i.e., returnees in Japanese). As such, this statistical difference is most likely driven by at most a few English monolingual children supplying lower amounts of NPs in this specific isolated performance.

In the Maintenance context, returnees in English produced more NPs than English and Japanese monolinguals, and returnees in Japanese produced more NPs than Japanese and English monolinguals, but no significant differences were found between English and Japanese in returnees as well as between English and Japanese monolinguals. That is, returnee children were redundant in both of their languages compared to their monolingual peers. Our finding adds to previous literature (Serratrice, 2007; Torregrossa et al., 2021; Torregrossa & Bongartz, 2018)—which only showed such overspecification behavior in one language—by demonstrating that the redundant use of NPs can manifest as a universal strategy that children apply to both languages.

The crucial question here is *why* bilingual children display such behavior. Torregrossa et al. (2021) attributes overspecification to “be the expected automatic outcome if—under reduced degrees of language experience—the proceduralization of the grammatical options for reference available in the language is not in place” (p.703). That is, children may not have fully acquired the grammatical function of pronominal forms since their use requires encoding of several grammatical distinctions such as number, gender, and case (in the case of overt pronouns) as well as planning at the global discourse level (such as accessibility and cohesion) (Ryan, 2015). Thus, they tend to opt for the most simple and available form that can alleviate processing load (Gullberg, 2006). Alternatively, Ryan (2015) explains overspecification as a clarity-based communication strategy, which bilinguals adopt as a cautious approach to avoid communicative breakdowns with the interlocutor. These accounts together can also explain why returnees produced more NPs in English than Japanese (and also more NPs compared to both monolingual groups) in the Reintroduction context where they are required to use explicit NPs to avoid ambiguity on the part of the listener. Although these returnee children were tested a few months after leaving an English-dominant environment, they are all indeed L2 speakers of English and have only lived in an English-dominant environment for 3.77 years on average. Furthermore, they received around 50% of their exposure in Japanese (mainly in the home context) even while residing abroad. Therefore, it is not too surprising that in topic-shift contexts where the stakes are higher for communication breakdown, bilingual children tend to be more explicit in the language that they are less experienced in and avoid the use of L2 English overt pronouns that can be more prone to errors (Serratrice, 2007).

In terms of the longitudinal results, we observed no changes in the returnees’ English NP use across the three time points tested in all referential contexts. In other words, we found no effects of L2 attrition, but rather a stable maintenance of their referential strategies over the course of five years. This seems to contrast with findings by Flores (2012) who showed effects of attrition in German object expression and verb placement among Portuguese-German returnees who moved back to Portugal before puberty (age 11/12). We must emphasize that since Flores’s (2012) study was conducted cross-sectionally and their linguistic performance was compared to bilingual child control groups still living in Germany, it is not entirely certain whether the non-target like performance observed in these returnees is evidence of regression or a result of arrested development. Although the returnees in our study showed different baseline performance to their monolingual peers, they clearly demonstrate that they are able to use referential forms appropriately in each context in both their L1 Japanese and L2 English, even with regards to particle use in Japanese and nominal markers in English.

By adopting a longitudinal paradigm, we can be more confident that the lack of change in referential strategies in L2 English is not due to arrested development of the structure. Furthermore, another factor that may contribute to these differences is the fact that despite having drastically reduced exposure to English (when compared to living abroad), Japanese returnees may have had more opportunities to engage in English via lessons at school, extra-curricular English activities via JOES, and social media (e.g., YouTube, movies, TV shows). Additionally, since English is a prestigious and widely studied language in Japan, the motivation for Japanese families to maintain their children’s English competence may be higher than it is for German maintenance in Portugal. After all, German is not an easily accessible or prioritized for studying language in Portugal (Flores & Kubota, 2023). Such socio-linguistic factors may affect the degree of attrition observed among different returnee populations, but more studies that attempt to quantify and integrate motivation, language preference, and family language policy factors in their empirical work are needed to properly explore these inquiries.

In contrast to the longitudinal L2 English results, for L1 Japanese, we found no changes in the proportion of NPs for First Mention, but there was a *decrease* in NP use from second to third session in the Maintenance context and an *increase* in NP use from first to third session and second to third session in the Reintroduction context. This pattern suggests that children were less likely to be redundant when maintaining a referent and used more appropriate explicit referent (full NP) when reintroducing a character, indicating improvement in referential strategies over time. In terms of the Maintenance context, interestingly, we see a U-shape-like acquisition curve in which the NP use increases (albeit insignificant) from first to second session and then decreases from second to third session. That is, they appear to initially apply the overspecification strategy in the Maintenance context to Japanese as a response to suddenly being immersed in an environment where they are constantly exposed to Japanese; opting for the most simple and available form (NPs) that can alleviate processing load. However, once they are immersed in the Japanese environment for several years, it is not costly anymore to process Japanese due to prolonged and increased exposure, and thus their use of NPs decreases from second to third session, using more context-appropriate null pronouns to maintain the referent.

Our findings differ from Antonova-Unlu et al. (2021) and Kaya-Soykan et al. (2023) where both studies found Turkish-German bilinguals who returned to Turkey (and have been living there for more than ten years) to diverge from their Turkish monolinguals counterparts in terms of their comprehension and production of evidentiality markers—a structure that is licensed by discourse information. Their findings suggest that re-immersion in the L1 environment did not re-activate or allow for reacquisition of evidentiality markers, while our study observed development/re-exposure effects of referential production over the course of five years since their return to the homeland. Such discrepancies may be attributed to age (of return) differences; the returnees in Antonova-Unlu et al. (2021) and Kaya-Soykan et al. (2023) left Germany after puberty while the Japanese-English bilinguals in our study all returned to Japan before puberty. The critical role that age plays in explaining re-exposure/relearning patterns have been discussed extensively in the literature mostly in regards to maturational vs. entrenchment accounts of age effects (see Bylund, 2009 for detailed discussion). Regardless of what specifically underlies the significance of age in bilingual re-exposure/relearning, it seems to be clear that children who are still well within the developmental trajectories (as in the returnees in the current study) are more susceptible to change in their linguistic system than adults who may have already stabilized their grammatical system at the point of return. Alternatively, it is possible that age is not critical here or it is interacting with a third factor related to the domains of grammar not being fully comparable (which cannot be teased apart *a posteriori*).It should not be ignored that reference and its resolution is universal—all languages must contend with it within the confines of their grammatical system—whereas evidentiality is something that languages vary on, that is, only some encode it grammatically. As a result, it might simply be the case that evidentiality is less likely to be affected by re-immersion period.

Crucially, the correlational analyses at individual time points and across time revealed a similar picture; there was a significant correlation between NP use and L2 exposure in the English Maintenance context at second session, indicating that returnee children who received *less* exposure in English after a year of return to Japan were *more* redundant when maintaining reference. Moreover, we also found a significant positive correlation between the difference score in exposure at Interval 3 and difference in proportions of NPs at Interval 3 for the English Maintenance context. That is, children who experienced *more loss* in English exposure since their return to Japan used *more redundant NPs* in the English maintenance context after five years of return to the homeland. These results together highlight the importance of language exposure in predicting overspecification behavior among bilingual children, adding support to the findings of Torregrossa et al. (2021) and Torregrossa & Bongartz (2018). Our results partially mirror Torregrossa and colleagues’ in terms of the lack of correlation between WM/updating and referential production, as they only found a relationship between N-back performance and redundant NP use in children who are dominant in the target language (N = 45), but not for the rest of the participants who were dominant in the non-target language (N = 80). Torregrossa et al. (2021) explains the absence of correlation in the latter group as follows: “the effects of EF-skills are overshadowed by the effects of language exposure, given that dominant experience in the non-target language leads to the same outcome in terms of reference use as low EF-skills, i.e., production of overspecificed full nouns” (p.704). Following this line of argument, in the case of returnee children in which *exposure* is the main variable that changes drastically as a consequence of environmental transitions, it is not surprising that exposure plays a greater role in predicting referential production than WM (although the effects of language experience and executive control or WM may interact, as alluded in Serratrice & De Cat, 2020).

**5. Conclusion**

The findings of our study highlight the differential trajectories of L1 re-exposure and L2 attrition/maintenance process among Japanese-English bilingual children. While their referential production in their L1 develops and improves over time—especially in contexts where bilinguals tend to be either overexplicit or under-informative—their L2 referential strategies did not undergo any changes, indicating stable maintenance of the structure. However, correlational analyses revealed that children who received less exposure in L2 English (at single time points and across time) showed more L2 overspecification behavior, underscoring the crucial role that continued language exposure plays in child bilingual attrition. Crucially, one can only appreciate the above observations fully by virtue of the longitudinal nature of the present design.

**References**

Antonova-Unlu, E., Wei, L., & Kaya-Soykan, D. (2021). Interfaces in the returnees’ heritage language: Is the complete (re-)activation possible? *International Journal of Bilingualism*, *25*(6), 1764–1781. https://doi.org/10.1177/13670069211043572

Ariel, M. (1996). Referring expressions and the+/-coreference distinction. *Pragmatics and beyond New Series*, 13–36.

Arizmendi, G. D., Alt, M., Gray, S., Hogan, T. P., Green, S., & Cowan, N. (2018). Do Bilingual Children Have an Executive Function Advantage? Results From Inhibition, Shifting, and Updating Tasks. *Language, Speech, and Hearing Services in Schools*, *49*(3), 356–378. https://doi.org/10.1044/2018\_LSHSS-17-0107

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, *25*(4), 657–689. https://doi.org/10.1007/s11049-007-9026-9

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press. https://books.google.com/books?hl=en&lr=&id=UGwwQ2Edo6gC&oi=fnd&pg=PP1&dq=Berman+and+Slobin+1994+referential&ots=fyJAn1vN9m&sig=NRlJ9xlJsuYw-fRuKVmLJ01ZSRY

Bylund, E. (2009). Maturational constraints and first language attrition. *Language Learning*, *59*(3), 687–715.

Chevalier, N. (2018). Willing to think hard? The subjective value of cognitive effort in children. *Child Development*, *89*(4), 1283–1295. https://doi.org/10.1111/cdev.12805

Chondrogianni, V., Vasić, N., Marinis, T., & Blom, E. (2015). Production and on-line comprehension of definiteness in English and Dutch by monolingual and sequential bilingual children. *Second Language Research*, *31*(3), 309–341. https://doi.org/10.1177/0267658314564461

Clancy, P. M. (1980). Referential choice in English and Japanese narrative discourse. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, *3*, 127–201.

Clancy, P. M. (1992). Referential strategies in the narratives of Japanese children. *Discourse Processes*, *15*(4), 441–467. https://doi.org/10.1080/01638539209544822

Colozzo, P., & Whitely, C. (2015). The interplay of referential function and character primacy on referring expressions in children’s narratives. *First Language*, *35*(2), 137–162. https://doi.org/10.1177/0142723715577321

De Houwer, A. (2021). *Bilingual development in childhood*. Cambridge University Press. https://www.cambridge.org/core/elements/bilingual-development-in-childhood/3517C7DA9EE432C32C3DCDB5AE18C375

Filiaci, F., Sorace, A., & Carreiras, M. (2014). Anaphoric biases of null and overt subjects in Italian and Spanish: A cross-linguistic comparison. *Language, Cognition and Neuroscience*, *29*(7), 825–843. https://doi.org/10.1080/01690965.2013.801502

Flores, C. (2010). The effect of age on language attrition: Evidence from bilingual returnees. *Bilingualism: Language and Cognition*, *13*(4), 533–546.

Flores, C. (2012). Differential effects of language attrition in the domains of verb placement and object expression. *Bilingualism: Language and Cognition*, *15*(3), 550–567.

Flores, C., & Kubota, M. (2023). *The effects of environment change on third languages: The case of returnees.*

Gangopadhyay, I., Davidson, M. M., Weismer, S. E., & Kaushanskaya, M. (2016). The role of nonverbal working memory in morphosyntactic processing by school-aged monolingual and bilingual children. *Journal of Experimental Child Psychology*, *142*, 171–194.

Gullberg, M. (2006). Handling Discourse: Gestures, Reference Tracking, and Communication Strategies in Early L2. *Language Learning*, *56*(1), 155–196. https://doi.org/10.1111/j.0023-8333.2006.00344.x

Hansen, L. B., Macizo, P., Duñabeitia, J. A., Saldaña, D., Carreiras, M., Fuentes, L. J., & Bajo, M. T. (2016). Emergent Bilingualism and Working Memory Development in School Aged Children. *Language Learning*, *66*(S2), 51–75. https://doi.org/10.1111/lang.12170

Hendriks, P. (2016). Cognitive modeling of individual variation in reference production and comprehension. *Frontiers in Psychology*, *7*, 506.

Hendriks, P., Koster, C., & Hoeks, J. C. J. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience*, *29*(4), 391–407. https://doi.org/10.1080/01690965.2013.766356

Kaya-Soykan, D., Antonova-Unlu, E., & Sagin-Simsek, C. (2023). The production and perception of Turkish evidentiality markers by Turkish-German returnees. *Applied Linguistics Review*, *14*(2), 251–270. https://doi.org/10.1515/applirev-2020-0042

Keijzer, M., & de Bot, C. (2018). Unlearning and relearning of languages from childhood to late adulthood. In *The Cambridge handbook of bilingualism* (pp. 267–286). Cambridge University Press. https://research.rug.nl/files/79350500/Ch15\_Keijzer\_deBot\_revised\_Feb\_16\_2017.pdf

Kubota, M., Chevalier, N., & Sorace, A. (2020). Losing access to the second language and its effect on executive function development in childhood: The case of’returnees’. *Journal of Neurolinguistics*, *55*, 100906.

Kuijper, S. J., Hartman, C. A., & Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PloS One*, *10*(7), e0132408.

Loban, W. (1976). *Language Development: Kindergarten through Grade Twelve. Research Report No. 18.* National Council of Teachers of English.

Lozano, C. (2006). The development of the syntax-discourse interface. In V. Torrens & L. Escobar (Eds.), *The acquisition of syntax in Romance languages* (pp. 371–399). John Benjamins.

Mayer, M. (1969). *Frog, where are you?* Dial Press.

Mayer, M. (1973). *Frog on his own*. Dial Press.

Mishina-Mori, S., Nagai, Y., & Yujobo, Y. J. (2018). Cross-linguistic influence in the use of referring expressions in school-age Japanese–English bilinguals. *Proceedings of the 42nd Annual Boston University Conference on Language Development*, 546–557. http://www.lingref.com/bucld/42/BUCLD42-42.pdf

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793–817.

Paradis, J., & Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language*, *30*(2), 371–393.

Ryan, J. (2015). Overexplicit Referent Tracking in L2 English: Strategy, Avoidance, or Myth? *Language Learning*, *65*(4), 824–859. https://doi.org/10.1111/lang.12139

Serratrice, L. (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English–Italian bilingual children. *Bilingualism: Language and Cognition*, *10*(3), 225–238.

Serratrice, L., & De Cat, C. (2020). Individual differences in the production of referential expressions: The effect of language proficiency, language exposure and executive function in bilingual and monolingual children. *Bilingualism: Language and Cognition*, *23*(2), 371–386.

Serratrice, L., & Hervé, C. (2015). Referential expressions in bilingual acquisition. In *The acquisition of reference* (pp. 311–333). John Benjamins. https://www.jbe-platform.com/content/books/9789027267894-tilar.15.13ser

Serratrice, L., Sorace, A., Filiaci, F., & Baldo, M. (2012). Pronominal objects in English–Italian and Spanish–Italian bilingual children. *Applied Psycholinguistics*, *33*(4), 725–751.

Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntax–pragmatics interface: Subjects and objects in English–Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, *7*(3), 183–205.

Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33.

Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguistic Approaches to Bilingualism*, *6*(5), 669–684.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339–368.

Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, *119*(3), 460–477.

Torregrossa, J., Andreou, M., Bongartz, C., & Tsimpli, I. M. (2021). Bilingual acquisition of reference: The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition*, *24*(4), 694–706.

Torregrossa, J., & Bongartz, C. (2018). Teasing apart the effects of dominance, transfer, and processing in reference production by German–Italian bilingual adolescents. *Languages*, *3*(3), 36.

Torregrossa, J., & Bongartz, C. (2023). Activation of referents in the bilingual mind. *Individual Differences in Anaphora Resolution: Language and Cognitive Effects*, *18*, 221.

Unsworth, S. (2016). Amount of exposure as a proxy for dominance in bilingual language acquisition. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 156–173). Cambridge University Press.

Vogels, J., Krahmer, E., & Maes, A. (2015). How Cognitive Load Influences Speakers’ Choice of Referring Expressions. *Cognitive Science*, *39*(6), 1396–1418. https://doi.org/10.1111/cogs.12205

Whitely, C., & Colozzo, P. (2013). Who’s Who? Memory Updating and Character Reference in Children’s Narratives. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1625–1636. https://doi.org/10.1044/1092-4388(2013/12-0176)

**Tables**

Table 1. *Proportion of language exposure to English (relative to Japanese) at first (Time 1), second (Time 2), and third (Time 3) test sessions*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Time 1 | Time 2 | Time 3 |  |
| **Mean** | 46.72 | 5.63 | 16.51 |  |
| **SD** | 13.38 | 6.32 | 12.27 |  |
| **Min** | 26.56 | 0 | 3.31 |  |
| **Max** | 82.45 | 20.50 | 51.72 |  |

Table 2. *Group pairwise comparisons of the proportions of NPs split by Context (First Mention, Maintenance, Reintroduction)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pairwise comparisons | estimate | SE | z ratio | p value |
| **First Mention** |  |  |  |  |
| eng\_mono - eng\_returnee | -1.28 | 0.45 | -2.84 | < .001 |
| eng\_mono - jap\_mono | -1.12 | 0.45 | -2.49 | 0.01 |
| eng\_mono - jap\_returnee | -1.11 | 0.48 | -2.33 | 0.02 |
| eng\_returnee - jap\_mono | 0.15 | 0.56 | 0.27 | 0.78 |
| eng\_returnee - jap\_returnee | 0.17 | 0.56 | 0.3 | 0.77 |
| jap\_mono - jap\_returnee | 0.01 | 0.58 | 0.02 | 0.98 |
| **Maintenance** |  |  |  |  |
| eng\_mono - eng\_returnee | -0.93 | 0.22 | -4.15 | < .001 |
| eng\_mono - jap\_mono | -0.01 | 0.24 | -0.05 | 0.96 |
| eng\_mono - jap\_returnee | -0.63 | 0.25 | -2.57 | 0.01 |
| eng\_returnee - jap\_mono | 0.91 | 0.23 | 3.96 | < .001 |
| eng\_returnee - jap\_returnee | 0.3 | 0.17 | 1.77 | 0.08 |
| jap\_mono - jap\_returnee | -0.62 | 0.25 | -2.45 | 0.01 |
| **Reintroduction** |  |  |  |  |
| eng\_mono - eng\_returnee | -0.62 | 0.2 | -3.03 | < .001 |
| eng\_mono - jap\_mono | 0.28 | 0.2 | 1.43 | 0.15 |
| eng\_mono - jap\_returnee | 0.17 | 0.2 | 0.87 | 0.39 |
| eng\_returnee - jap\_mono | 0.9 | 0.21 | 4.32 | < .001 |
| eng\_returnee - jap\_returnee | 0.79 | 0.13 | 6.21 | < .001 |
| jap\_mono - jap\_returnee | -0.11 | 0.2 | -0.54 | 0.59 |

Table 3. *Time pairwise comparisons of the proportions of NPs split by Language (Japanese, English) and Context (First Mention, Maintenance, Reintroduction)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pairwise comparisons | estimate | SE | z ratio | p value |
| **English First Mention** |  |  |  |  |
| T1 - T2 | -1.15 | 0.8 | -1.44 | 0.15 |
| T1 - T3 | 0.37 | 0.54 | 0.69 | 0.49 |
| T2 - T3 | 1.52 | 0.8 | 1.91 | 0.06 |
| **Japanese First Mention** |  |  |  |  |
| T1 - T2 | -0.03 | 0.59 | -0.05 | 0.96 |
| T1 - T3 | -1.22 | 0.81 | -1.5 | 0.13 |
| T2 - T3 | -1.19 | 0.82 | -1.46 | 0.14 |
| **English Maintenance** |  |  |  |  |
| T1 - T2 | -0.02 | 0.15 | -0.16 | 0.87 |
| T1 - T3 | 0.11 | 0.16 | 0.65 | 0.52 |
| T2 - T3 | 0.13 | 0.16 | 0.79 | 0.43 |
| **Japanese Maintenance** |  |  |  |  |
| T1 - T2 | -0.29 | 0.22 | -1.33 | 0.18 |
| T1 - T3 | 0.29 | 0.21 | 1.36 | 0.17 |
| T2 - T3 | 0.58 | 0.22 | 2.61 | 0.01 |
| **English Reintroduction** |  |  |  |  |
| T1 - T2 | 0.15 | 0.15 | 0.98 | 0.33 |
| T1 - T3 | -0.01 | 0.16 | -0.08 | 0.93 |
| T2 - T3 | -0.16 | 0.15 | -1.05 | 0.3 |
| **Japanese Reintroduction** |  |  |  |  |
| T1 - T2 | -0.09 | 0.14 | -0.65 | 0.52 |
| T1 - T3 | -0.44 | 0.15 | -2.88 | < .001 |
| T2 - T3 | -0.35 | 0.15 | -2.37 | 0.02 |

Table 4. *Pearson bivariate correlation matrix between relative language exposure (Exposure) and accuracy on the N-back task (WM) at Time 1 (T1), Time 2 (T2), Time 3 (T3) and proportions of NP use at Time 1, Time 2, Time 3 in each context (First Mention, Maintenance, Reintroduction) and language (English, Japanese). Values indicate Pearson’s r. Bolded values with an asterisk indicate p < .05.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **First Mention** | | | | | | |
|  | English NP use | | | Japanese NP use | | |
|  | T1 | T2 | T3 | T1 | T2 | T3 |
| Exposure (T1/T2/T3) | -0.33 | -0.32 | -0.38 | 0.34 | -0.06 | -0.37 |
| WM (T1/T2/T3) | -0.01 | 0.08 | -0.23 | -0.10 | 0.17 | -0.03 |
| **Maintenance** | | | | | | |
|  | English NP use | | | Japanese NP use | | |
|  | T1 | T2 | T3 | T1 | T2 | T3 |
| Exposure (T1/T2/T3) | -0.27 | **-0.48\*** | -0.38 | 0.34 | 0.12 | -0.04 |
| WM (T1/T2/T3) | 0.27 | 0.12 | -0.23 | -0.10 | -0.03 | -0.22 |
| **Reintroduction** | | | | | | |
|  | English NP use | | | Japanese NP use | | |
|  | T1 | T2 | T3 | T1 | T2 | T3 |
| Exposure (T1/T2/T3) | -0.18 | 0.00 | -0.19 | 0.19 | -0.19 | -0.14 |
| WM (T1/T2/T3) | 0.21 | -0.07 | -0.23 | 0.03 | 0.12 | 0.03 |

Table 5. *Pearson bivariate correlation matrix between different scores of (i) relative language exposure (Exposure) at Interval 1 (Time 2 – Time 1), Interval 2 (Time 3– Time 2), Interval 3 (Time 3 – Time 1) (ii)WM at Interval, Interval 2, Interval 3 and (iii) difference in proportions of NPs at Interval 1, Interval 2, Interval 3 in each context (First Mention, Maintenance, Reintroduction) and language (English, Japanese). Values indicate Pearson’s r. Bolded values with an asterisk indicate p < .05.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **First Mention** | | | | | | |
|  | English NP use | | | Japanese NP use | | |
|  | Interval 1 | Interval 2 | Interval 3 | Interval 1 | Interval 2 | Interval 3 |
| Difference score in Exposure  (Interval 1/2/3) | -0.13 | 0.24 | 0.25 | 0.11 | -0.03 | 0.05 |
| Difference score in WM (Interval 1/2/3) | -0.30 | 0.24 | 0.21 | -0.01 | 0.28 | -0.03 |
| **Maintenance** | | | | | | |
|  | English NP use | | | Japanese NP use | | |
|  | Interval 1 | Interval 2 | Interval 3 | Interval 1 | Interval 2 | Interval 3 |
| Difference score in Exposure  (Interval 1/2/3) | -0.03 | -0.04 | **-0.40\*** | -0.15 | 0.06 | -0.16 |
| Difference score in WM (Interval 1/2/3) | 0.15 | 0.05 | 0.18 | -0.03 | 0.05 | -0.31 |
| **Reintroduction** | | | | | | |
|  | English NP use | | | Japanese NP use | | |
|  | Interval 1 | Interval 2 | Interval 3 | Interval 1 | Interval 2 | Interval 3 |
| Difference score in Exposure  (Interval 1/2/3) | -0.09 | 0.22 | -0.24 | 0.05 | 0.06 | 0.10 |
| Difference score in WM (Interval 1/2/3) | -0.08 | -0.21 | 0.01 | -0.39 | -0.08 | 0.19 |

**Figures**

A graph of different colored squares

Description automatically generated

Figure 1. Proportions of different particle use (ga, null, wa) in First Mention (FM), Maintenance (M) and Reintroduction (R) contexts in Japanese among monolingual Japanese children and returnees in first, second, third test sessions.

A graph of different colored squares

Description automatically generated with medium confidence

Figure 2. Proportions of different nominal markers (definite, indefinite, pronominals) in First Mention (FM), Maintenance (M) and Reintroduction (R) contexts in English among monolingual English children and returnees in first, second, third test sessions.

A diagram of different colored boxes

Description automatically generated

Figure 3. Proportions of NPs produced within each (FM =First Mention, M = Maintenance, R = Reintroduction) for each language/group (Japanese returnee, Japanese monolinguals, English returnee, English monolinguals).

![A graph of different colored boxes

Description automatically generated with medium confidence]()

Figure 4. Proportions of NPs produced within each context (FM =First Mention, M = Maintenance, R = Reintroduction) for each Time (Time 1, Time 2, Time 3) and Language (English, Japanese) among the returnee children.

**Data Availability Statement**

The data that support the findings of this study are openly available in https://osf.io/qh3x4/