International Journal of Linguistics Studies
ISSN: 2754-2599
DOI: 10.32996/ijls
Journal Homepage: www.al-kindipublisher.com/index.php/ijlss

IJLS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

The Acquisition of Binding Principles across Languages: A Literature Review

Halima Alhabib
Department of English Language and Literature, College of Languages and Humanities, Qassim University, Buraydah 52571, Saudi Arabia
Corresponding Author: Halima Alhabib, **E-mail**: H.alhabib@qu.edu.sa

| ABSTRACT

The acquisition of Binding Theory (Chomsky, 1981) has been a central battlefield in generative child language acquisition and Second Language Acquisition for over three decades. Cross-linguistic evidence demonstrates early, adult-like mastery of Principles A and C, consistent with Universal Grammar. However, children's interpretation of Principle B has historically exhibited a pronounced delay—widely termed the Pronoun Interpretation Problem (PIP). This review provides a comprehensive synthesis of the theoretical and empirical literature surrounding the PIP. We highlight the traditional contrast between the acquisition patterns of the binding constraints, and link the relevant explanatory accounts, spanning pragmatic deficits (Principle P), initial governing category parameterization, working memory constraints on reference-set computation (Reinhart, 2006), and the Quantificational Asymmetry and Antecedent-Contained Deletion ellipsis (Chien & Wexler, 1990; Kiguchi & Thornton, 2004). Moreover, we provide a detailed evaluation of first language acquisition research in Arabic dialects. Finally, we review recent studies re-examining the persistence of the PIP under task-bias-controlled paradigms, including eye-tracking with toddlers (Emond & Shi, 2024), the satisfaction of pragmatic felicity conditions (Conroy et al., 2009), and the Coloring Book Task (Pinto & Zuckerman, 2019). We conclude that the PIP remains a vibrant area for investigation, situated at the critical intersection of contrasting standpoints.

| KEYWORDS

Binding Theory, Principle B Delay, Pronoun Interpretation Problem (PIP), Task Artifacts, Arabic Language Acquisition

| ARTICLE INFORMATION

ACCEPTED: 01 October 2026 **PUBLISHED:** 02 November 2026 **DOI:** 10.32996/ijls.2026.6.4.3

1. Introduction

In the field of language acquisition, whether First (L1A) or Second (SLA), the interpretive properties of pronominal and reflexive binding relations have generated extensive empirical experimentation and vigorous theoretical debate from the 1980s to the present (Al Kafri, 2019; Chien & Wexler, 1990; Conroy et al., 2009; Emond & Shi, 2024; Grolla, 2005; Kiguchi & Thornton, 2004; Kim et al., 2015; Matsuo, 1999; McDaniel et al., 1990; McKee, 1992; Sanoudaki & Varlokosta, 2015; Slabakova et al., 2017; Su, 2020; White, 1998). Beyond evaluating each principle in isolation, researchers have systematically compared developmental trajectories across all three binding constraints, i.e., Binding Principles A, B, and C. A broad consensus has emerged that young children demonstrate intact knowledge of Principles A and C from early preschool age (Matsuo, 1999; Su, 2020). In contrast, studies on the acquisition of Principle B across multiple languages reported a pronounced delay in children's mastery of pronominal non-coreference—widely referred to as the Pronoun Interpretation Problem (PIP) or Delay of Principle B Effect (DPBE) (Al Kafri, 2019; Chien & Wexler, 1990; McDaniel et al., 1990; Ruigendijk et al., 2010; Slabakova et al., 2017). Similarly, within the Arabic linguistic domain, experimental acquisition studies (Bolotin, 2002; Mustafawi & Mahfoudhi, 2005) reported that child performance on Principle B lagged behind Principle A.

Crucially, findings surrounding the PIP were not predominantly interpreted as a refutation of the innate status of the Principle B constraint itself. Instead, a diverse range of theoretical accounts was formulated to explain why children accept illegitimate local coreference. These accounts span developmental pragmatic deficits (e.g., the pragmatic Principle P), initial governing category

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

parameterization, computational complexity and working memory constraints, and accidental coreference strategies. To verify that the subjects had knowledge of Principle B, several studies tested pronouns in constructions beyond referential antecedents, such as Antecedent-Contained Deletion (ACD) constructions (Kiguchi & Thornton, 2004) and those with quantified antecedents (Chien & Wexler, 1990; McDaniel et al., 1990; Slabakova et al., 2017). These studies reported significant improvements in children's performance, suggesting that they indeed possess knowledge of Principle B. More recently, however, several researchers have argued that the Quantificational Asymmetry (QA) observed in earlier studies, as well as the PIP itself, was largely an artifact of flawed experimental designs and heavy task demands (Conroy et al., 2009; Elbourne, 2005; Emond & Shi, 2024; Pinto & Zuckerman, 2019). When tested under low-demand, pragmatically felicitous experimental conditions, children demonstrate intact adherence to Principle B without exhibiting the PIP (Conroy et al., 2009; Emond & Shi, 2024; Pinto & Zuckerman, 2019; Sanoudaki & Varlokosta, 2015).

To bridge these theoretical and methodological divides, the primary objective of this paper is to review the competing syntactic, pragmatic, processing, and task-induced accounts of the PIP. Rather than advancing a single, overarching explanation, it emphasizes that the acquisition of pronominal binding remains a multifaceted phenomenon with varying trajectories across different languages, generating fruitful empirical questions surrounding cross-linguistic variation.

The remaining sections of this paper are organized as follows: Section 2 establishes the theoretical foundations of Binding Theory within the framework of Universal Grammar. Section 3 surveys the empirical trajectories of pronominal binding across both L1A and SLA. Section 4 focuses on the acquisition of binding principles in spoken Arabic dialects. Section 5 investigates key structural exceptions to the PIP, specifically detailing bound variable pronouns under the QA paradigm and ellipsis in ACD constructions. Section 6 evaluates the primary theoretical accounts proposed to explain the PIP delay, contrasting syntactic, pragmatic, processing, and task-induced explanations. Finally, Section 7 provides a summary of the literature and outlines future directions for research on Binding Theory acquisition.

2. Binding Theory and Universal Grammar

As a core module of generative grammar, Binding Theory establishes the formal structural conditions governing the referential dependencies between reflexives, pronouns, and referential expressions (R-expressions) and their potential antecedents within a clause (Chomsky, 1981). The formal operations of Binding Theory rely on three foundational structural concepts: coindexation, c-command, and the governing category (or local domain). Coindexation (represented by identical subscripts, e.g., $_i$) denotes referential identity between nominals. Syntactic prominence is defined by c-command (Reinhart, 1976), wherein a node α c-commands a node β if neither dominates the other and the first branching node dominating α also dominates β. An expression α is syntactically bound by β if and only if β c-commands α and α and β are coindexed; if α is not bound, it is syntactically free (Chomsky, 1981). Finally, the governing category represents the minimal structural domain (typically the minimal IP/TP) containing the nominal, its lexical governor, and an accessible subject (Chomsky, 1981). Chomsky (1981) formulates three universal principles:

- Principle A: An anaphor must be bound in its governing category.
- Principle B: A pronominal must be free in its governing category.
- Principle C: An R-expression must be free everywhere.

Under this formulation, Principles A and B predict a strict distributional contrast. As Guasti (2002) notes, Principles A and B "imply a strict complementarity between reflexives and pronouns: by and large a reflexive must be bound in those environments in which a pronoun must be free" (p. 285). This structural complementarity is illustrated in (1), where the reflexive *himself* in object position must be coindexed with the local c-commanding subject *Kim*, whereas the pronominal *him* is prohibited from coreferring with *Kim*:

(1) Kim$_i$ hit himself$_i$ / him*$_i$/$_j$.

In child language acquisition, the empirical study of these principles offers insight into how Universal Grammar (UG) shapes language development across typologically diverse human languages. While core binding constraints are hypothesized to be universal, languages exhibit cross-linguistic variations in domain size, morphological proform types (e.g., clitics vs. full proforms), and long-distance binding possibilities. From a parameter-setting perspective, Chomsky (1981) notes: "The language learner equipped with the theory of UG as a part of the initial state requires evidence to fix the parameter and then knows the other properties of the language that follow from this choice of value" (p. 241). While Binding Theory provides the formal structural architecture governing anaphoric distribution, empirical acquisition research investigates how children operationalize this innate

knowledge during real-time sentence processing, leading to the central empirical debate surrounding the contrastive acquisition of Principles A, C, and B.

3. Empirical Trajectories in L1A and L2A

3.1 First Language Acquisition (L1A)

In a classic study on the PIP, Chien and Wexler (1990) investigated English-speaking children aged 2;6 to 6;6 using traditional methods like act-out and yes/no judgment tasks. At least 120 children were individually tested in each experiment. The children demonstrated firm knowledge of Principle A between ages 5 and 6 (~90% accuracy), but they persistently accepted local coreference for pronouns in Principle B test sentences when gender cues were absent. Children exhibited minimal age-related improvement, committing frequent Principle B violations up to age 6;6. Likewise, McDaniel et al. (1990) found that English-speaking children aged 2 to 6 strictly abided by Principles A and C, but repeatedly violated Principle B in both act-out and judgment tasks, despite showing a preference for external referents for pronouns in the act-out task. Grolla (2005) similarly documented non-adult-like performance on pronominal judgment tasks among Brazilian Portuguese- and English-speaking children.

Cross-linguistic evidence, however, reveals striking typological variation. McKee (1992) conducted a comparative study of Italian-speaking and English-speaking children (aged 2;6 to 5;5) using a Truth Value Judgment Task (TVJT). While both groups displayed high accuracy on reflexive binding (95% for Italian; 91% for English), their performance on Principle B diverged radically: Italian children achieved 90% accuracy with object clitic pronouns, whereas English children scored only 61% with full object pronouns. The results of the English-speaking children contrasted with those of the Italian children but mimicked earlier English findings (e.g., Chien & Wexler, 1990; McDaniel et al., 1990).

Regarding Principle A sub-types, Matsuo (1999) examined the comprehension of reciprocal anaphors *each other* among 16 English-speaking children aged 3;11 to 5;10. Matsuo found that the children were able to interpret reciprocal sentences correctly in simple monoclausal sentences that included only one potential antecedent, but experienced difficulty in interpreting more complex sentences that involved multiple potential antecedents. Turning to Principle C, Su (2020) explored the acquisition of Principle C by Mandarin preschool children and found that they interpreted this principle correctly as early as 4 years old, confirming that structural c-command constraints on R-expressions are operational from early childhood.

3.2 Second Language Acquisition (SLA)

White (1998) observed that the delay in Principle B in L1 and L2 children's acquisition is often attributed to extra-grammatical factors that disappear when children grow up. In response, she hypothesized that the adults would not show the same problems with Principle B as children. Testing adult L2 learners of English, White found that intermediate L2 learners of English know that pronouns bind with antecedents outside the local domain. Based on this finding, White (1998) called for future research to focus on lower-level adult learners and predicted "that low-proficiency learners will perform similarly to the adult intermediate subjects, rather than to child L1 or L2 acquirers" (p. 435).

To empirically test this prediction, Slabakova et al. (2017) explored the PIP among 125 adult French and Spanish L1 speakers learning English as an L2 to investigate whether L2 learners would encounter problems with pronoun reference, similar to child first language acquisition. Their results revealed that the L2 learners showed the lowest accuracy when using full pronouns, rather than reduced pronouns (such as *'m*, for him), bound with referential antecedents. In the case of proficiency levels, the performance of L2 learners improved significantly as their proficiency test scores increased. Nonetheless, the learners with lower proficiency displayed a similar PIP pattern (but less severe) compared to previous research results of child L1 acquirers, despite diverse differences in circumstances such as age and cognitive abilities, contradicting White's (1998) earlier expectation.

Moreover, Kim et al. (2015) explored the interpretation of reflexives and pronouns among Korean native adult L2 learners of English using an eye-tracking experiment. They found that the participants generally possess knowledge of Principle A, but fail to have a native-like interpretation of Principle B. Similarly, Al Kafri's (2019) study compared the performance of adult and child L2 learners of English (Arabic and Chinese L1 speakers). The participants' overall responses showed that they were confused between reflexives and pronouns, considering most of the items as reflexives. Specifically, the child participants scored less than 70% on the pronoun sentences.

Providing perspective across age ranges of L2 exposure, Lee and Schachter (1997) examined the interpretation of English reflexives and pronouns among 76 Korean-English bilingual speakers, who were exposed to English for approximately the same time; 3 years. Their ages ranged from 6 to 24 years old, and they were divided into five groups according to their age: 6-7, 8-10, 11-13, 14-16, and 17-24 years old. They hypothesized that while three years of exposure provides a sufficient timeframe for

acquisition, both very young participants (who had not yet reached peak sensitivity) and older participants (who had passed the sensitive window) would fail to display native-like performance.

Their results showed that speakers aged 6-7 years were on the verge of recognizing Principle A. However, there was little evidence of Principle B recognition for this age group. The 8-10 age group showed stronger evidence for Principle A knowledge. As for Principle B, they did not reach native-like knowledge but were on an upward trend toward it. The speakers aged 11 to 13 demonstrated a consistent knowledge of both Principles A and B. The 14-16 age group revealed signs of declining ability to interpret Principle A, particularly with long-distance reflexives, and they were considerably less than native-like in Principle B. The 17-24 age group showed more declining ability in both Principles A and B. In sum, both the younger and older groups failed to reach a native-like performance in Principles A and B.

Taken together, these L2 investigations demonstrate that while L2 learners across age groups display robust knowledge of Principle A, Principle B remains uniquely vulnerable to illegitimate local coreference.

4. The Acquisition of Binding Principles in Arabic Dialects

Empirical investigations within Arabic dialects remain relatively limited. To date, only two studies, to my knowledge, have assessed Arabic-speaking children's knowledge of Principles A and B (Bolotin, 2002; Mustafawi & Mahfoudhi, 2005).

In an early investigation into the acquisition of Najdi Arabic, Bolotin (2002) conducted an experiment on 12 Saudi Arabic-speaking children aged between 5 and 13 to test their knowledge of the Binding Principles A and B. She found that the children demonstrated greater mastery of Principle A than Principle B. The test included 15 sentences: five sentences on Principle A, five on Principle B, and five ambiguous sentences, as illustrated in (2), (3), and (4), respectively:

(2) *Fariid gaal ʔin Hasan sara kitaab linafsah.*
Farid said that Hasan bought book-for himself
'Farid said that Hasan bought a book for himself.'

(3) *ʕaziiz gaal ʔin nabil šara kitaab lah.*
Aziz said that Nabil bought book for-him.
'Aziz said that Nabil bought a book for him.'

(4) *najiib wara husayn surtah.*
Najib showed Hussein his-picture
'Najib showed Hussein his picture.' (Bolotin, 2002, p. 214-215)

Bolotin reported that the children had a 92% accuracy rate on principle A, whereas on principle B, they had a mere 35%. Specifically, "on the Principle B sentences, four children got only one of the five sentences correct, seven got two right, and only one child got three right" (Bolotin, 2002, p. 215). Additionally, it was found that there was no improvement for Principle B with age. Finally, the paper concludes that "Arabic-speaking children appear to acquire command of Principle B later than children with other L1" (Bolotin, 2002, p. 218).

However, the findings of this study should be interpreted with caution, especially regarding the low performance on Principle B, due to the wide age range, limited number of participants, absence of an adult control group, and the metalinguistic nature of the test (which required children to explicitly identify referents in isolated sentences without visual or narrative contexts).

Focusing on Qatari Arabic (Q-Arabic), Mustafawi and Mahfoudhi (2005) investigated whether there were differences in children's acquisition of reflexive proforms and those of inherently reflexive verbs, and whether Q-Arabic clitic pronouns would accelerate acquisition, as found with clitic Italian pronouns (McKee, 1992). The participants included 30 children of equal gender distribution, divided into three age groups: 10 children aged between 3;3 and 3;11, 10 children aged between 4;2 and 4;11, and 10 children aged between 5;4 and 5;1. The experiment test was a TVJT comprising 40 test sentences: 16 with reflexive proforms, 16 with clitic pronouns attached to the verb, and eight with inherently reflexive verbs. The results indicated that all child groups displayed knowledge of Principle A, with no significant difference in the acquisition between verbs with reflexive proforms and inherently reflexive verbs. However, the children performed better in Principle A than in Principle B. The children's performance in the mismatch condition improved with age, although not statistically significant. The researchers deduced that the reason for the children's initial poor performance in the mismatch condition was due to the general cognitive development of children.

Furthermore, their results revealed that cliticization in Q-Arabic did not accelerate Principle B acquisition, unlike Italian clitics (McKee, 1992). Mustafawi and Mahfoudhi attributed this to the fact that Q-Arabic clitics occupy the position of an NP inside the

VP, thereby occupying a larger binding domain than Italian clitics, which are affixed to the verbs in the inflectional domain (INFL) and thus have a smaller binding domain. Moreover, clitics in Q-Arabic can occur in ambiguous readings since they can sometimes be used co-referentially, as illustrated in (5), making them more challenging to master than anaphors.

(5) *Ahmed šaraa l-a ktaab*.
Ahmed bought for-him book
'Ahmed bought a book for himself/him.' (Mustafawi & Mahfoudhi, 2005, p. 37)

Spoken Arabic dialects exhibit rich morphosyntactic variation, including pronominal cliticization, pro-drop properties, and specific locality conditions governing prepositional phrases, providing a promising domain of investigation for Binding Theory.

5. The Quantificational Asymmetry (QA) and Ellipsis

5.1 Bound Variable Pronouns

In an endeavor to determine whether the PIP reflects genuine syntactic deficit, researchers introduced the QA paradigm, contrasting referential antecedents (e.g., *Mama Bear*) with quantified antecedents (e.g., *every bear*). For instance, in their fourth experiment, Chien and Wexler (1990) examined whether children recognize Principle B in constructions that included bound variables by employing explicit quantifiers like '*every*' preceding NP antecedents, as in (6), in comparison to pronouns preceded by referential antecedents.

(6) Is every bear$_i$ touching her*$_i$/$_j$? (Chien & Wexler, 1990, p. 263)

The experimental question types used were name-reflexive, name-pronoun, quantifier-reflexive, and quantifier-pronoun questions. As a result, the children showed a divergence in comprehension between Principle A and Principle B under referential antecedents, reflecting a clear PIP. However, in quantifier trials, they demonstrated knowledge that pronouns cannot bind with a local quantified NP, suggesting that this discrepancy between these two principles does not necessarily mean that the children do not have knowledge of Principle B. Similarly, McDaniel et al. (1990) tested whether children would obey Principle B in bound variable contexts like (7):

(7) Everyone$_i$ is patting him*$_i$/$_j$. (McDaniel et al., 1990, p. 133)

Consequently, their results conformed with Chien and Wexler (1990), revealing that children obey Principle B only when the antecedent is quantified. In Contrast, Grolla (2005) reported a non-replication in Brazilian Portuguese, where children's performance with pronouns preceded by either a quantified or referential antecedent was around chance level (50% correct responses).

In a critical review of previous experimental studies on Principle B acquisition, Elbourne (2005) argued against the validity of the previously reported QA in children's pronoun interpretation, claiming that it may be the result of methodological flaws in experimental design, such as referent salience. However, in more recent SLA research that ensured the satisfaction of task assumptions as raised in previous work on Principle B acquisition, including Conroy et al. (2009), Crain and Thornton (1998), and Elbourne (2005), Slabakova et al. (2017) found that L2 learners were more accurate when using quantified antecedents compared to referential antecedents, thus aligning with Chien and Wexler's (1990) account of bound pronouns. Overall, findings from the QA paradigm demonstrate that language acquirers possess intact syntactic knowledge of Principle B when pronouns function as bound variables, suggesting that the underlying causes of the PIP lie somewhere other than a deficit in innate knowledge.

5.2 Ellipsis and Antecedent-Contained Deletion (ACD)

Kiguchi and Thornton (2004) conducted an experiment to investigate whether 4- and 5-year-old children interpret pronouns in ACD constructions, specifically in structures relevant to Binding Principles B and C. A typical ACD construction is exemplified in (8), while (9) illustrates ACD with Principle B:

(8) Kermit found Emily the same color egg that Gonzo did.
(9) The Mermaid baked him$_j$ the same food that Cookie Monster$_k$ did.
(10) [the same food that Cookie Monster$_k$ <baked him$_j$>]$_1$ the Mermaid baked him$_j$ t_1.
(11) Dora gave him the same color paint the Smurf's father did.
(12) Gumby brushed him. (Kiguchi and Thornton, 2004, pp. 252–257)

Kiguchi and Thornton assumed that the deleted VP is recovered at the level of Logical Form (LF) after Quantifier Raising (QR) occurs. Consequently, as shown in the LF representation in (10), the NP *Cookie Monster* c-commands the pronoun *him* inside the elided VP within the same clause, prohibiting an anaphoric reading under Principle B.

To test this, they used a TVJT in which children judged whether a puppet's statement at the end of a story was true or false. Target ACD sentences like (9) were constructed to be false under the adult reading to satisfy the assumption about children's tendency to judge sentences as true whenever possible, thus ensuring that rejections reflected true grammatical constraints. Additionally, to ensure that the children rejected the sentences for the right reasons, the children were asked to clarify their rejection and say what truly happened in the story.

The experiment comprised two sessions. Session 1 presented children with eight sentences: four target ACD sentences like (9) and four non-binding ACD controls like (8). Session 2 also contained eight sentences: four ACD binding controls like (11) and four matrix Principle B controls like (12). The non-binding ACD controls (8) verified whether children could interpret ACD ellipsis in isolation from binding principles. Meanwhile, the ACD binding controls (11) differed from (9) in that the name inside the relative clause was embedded within a possessive phrase (*the Smurf's father*). If children applied Principle C to the surface form, they were expected to incorrectly reject (11); conversely, if they applied Principle C at LF (where the possessive phrase prevents c-command), they would correctly accept it. Finally, the simple matrix sentences (12) established a baseline for children's performance on Principle B with referential antecedents.

The findings revealed that children accurately rejected the target ACD sentences (9) 85% of the time, while successfully accepting the ACD controls (8) at 97.7% and the ACD binding controls (11) at 93.3%. These results demonstrated that children's responses were adult-like: they exhibited intact knowledge of Principle B within elided VP structures and correctly applied Principle C at LF after QR. Furthermore, children demonstrated firm knowledge of Principle B with referential pronouns. Comparing the acceptance rate of ACD target sentences (15%) to that of matrix Principle B controls (27%) revealed that children permitted illicit coreference more frequently in simple matrix sentences than in ACD constructions. Thus, children proved more accurate at interpreting pronouns in computationally complex ACD sentences than in simple matrix sentences.

This research contributed what the authors termed a 'purer' test of Principle B to the binding literature. In ACD constructions, unlike in standard matrix sentences, there is no ambiguity between a bound variable and a referential pronoun, thereby eliminating potential local coreference readings. "The finding is, indeed, that children are more accurate at rejecting Principle B violations in ACD sentences (with a referential NP antecedent) than has been shown in previous experiments" (Kiguchi and Thornton, 2004, p. 266).

6. Theoretical Accounts of the Pronoun Interpretation Problem

Early scholars considered whether the PIP stemmed from incomplete lexical categorization, where children miscategorize personal pronouns as reflexives (Jakubowicz, 1984). Specifically, this view posited that children had not yet categorized pronouns as belonging to a distinct syntactic class from reflexives. McDaniel et al. (1990) explicitly dismissed this explanation; because children in their study successfully obeyed Principles A and C, they argued it was empirically untenable to maintain that children had categorized reflexives and R-expressions while failing to categorize personal pronouns. Consequently, McDaniel et al. concluded that children possess syntactic Principle B but lack mature pragmatic knowledge restricting pronominal coreference.

Chien and Wexler (1990) formulated the pragmatic Principle P account, arguing that syntactic Principle B (which governs bound variable binding) is intact in young children, but pragmatic Principle P (which governs pragmatic coreference between referential NPs) is delayed in development. They emphasized that the apparent developmental lag on Principle B sentences does not imply a syntactic deficit, but "rather, it is the pragmatic Principle P that is missing in young children" (Chien & Wexler, 1990, p. 274). Under adult grammar, Principle P blocks coreference between a pronoun and a local antecedent unless specific pragmatic context justifies it. Children's failure to apply Principle P leads them to accept accidental coreference in referential contexts while correctly rejecting local binding with quantified antecedents.

Adopting a purely syntactic perspective, Varela (1989) and McKee (1992) proposed that children initially assume a minimal governing category (e.g., VP) for personal pronouns: This proposal "allows children to obtain different domains for anaphors and pronouns, as well as for clitics and full proforms, with one formulation of governing category" (McKee, 1992, p. 48). Under this initial parameter setting, full VP-internal pronouns in English fall within a minimal domain that permits local coreference, generating the PIP. Conversely, Italian clitic pronouns move out of the VP to INFL, placing them outside the minimal domain and preventing coreference errors. Support for this comes from McKee's experiments, in which English-speaking children exhibited such non-adultlike coreference patterns, unlike Italian children, who had no problems with clitic pronouns.

In an influential series of papers, Reinhart (2006, 2011; Grodzinsky & Reinhart, 1993) rejected parameter mis-setting and syntactic ignorance accounts. Reinhart pointed out that if children genuinely lacked a syntactic rule or set a parameter incorrectly, experimental results should yield 90%–100% systematic non-adult errors. Instead, children in traditional studies consistently performed at chance level (~50%), showing an erratic guessing pattern within the same experimental condition. Reinhart proposed a processing-based account centered on reference-set computation. When evaluating a sentence with a referential antecedent (e.g., *Mama Bear is touching her*), the child's processor must construct two distinct syntactic representations—one with variable binding and one with coreference—and compare their semantic interpretations. This reference-set comparison places an exorbitant load on working memory. Because young children possess limited working memory capacity, their processing system experiences computational overload, forcing them to collapse into a guessing strategy (~50% performance). Grolla (2005) strongly support this processing-load model.

From a cross-linguistic perspective, Ruigendijk et al. (2010) found that German children correctly interpreted pronouns by the age of 4, whereas Hebrew children allowed local binding up to the age of 6. Ruigendijk et al. explained this difference through language-specific coreference availability. In German, children do not need to compute coreference against binding because local coreference is less available. However, in Hebrew, coreference is available; hence, children must process two options against each other, and their limited computation capacity interferes with the task.

A growing body of research has questioned the robustness of the PIP findings, suggesting that the reported results may be overstated or methodologically driven. In a landmark critique, Conroy et al. (2009) challenged the empirical foundation of the PIP and QA, arguing that the PIP may have been exaggerated by previous methodologies, particularly the use of TVJTs, rather than reflecting a deficit in grammatical representation. They reasoned that children's susceptibility to Principle B errors stems from pragmatic infelicities and the failure to implement task assumptions. That is, the traditional TVJTs failed to satisfy two mandatory pragmatic felicity conditions:

- *Availability*: The intended extra-sentential (deictic) referent must be made highly salient in the discourse context. If the deictic antecedent is not readily available, children are pragmatically driven to select the only prominent antecedent in the story; the local subject.
- *Disputability*: The false anaphoric interpretation must represent a plausible, disputable scenario within the context so that rejecting it is pragmatically motivated (the Condition of Plausible Dissent).

Across three experiments, Conroy et al. aimed to understand the vast variations in the findings of previous studies. Sixteen English-speaking children aged from 4;0 to 5;6 years were recruited in each experiment. The findings demonstrated that when *Availability* and *Disputability* were satisfied, children rejected local pronominal coreference with referential antecedents at adult-like levels, eliminating both the PIP and QA. Furthermore, when methodological flaws were deliberately reintroduced in Experiment 3, the PIP and QA reemerged, proving that the asymmetry was an artifact of experimental design. In the end, the researchers caution linguists against making hasty conclusions about the interaction between grammatical knowledge and linguistic behavior, stating that this relation is of high complexity. "Thus, in order to assess whether a behavior reflects grammatical structure, an experimenter must take great care to neutralize any possible influence from extragrammatical factors" (Conroy et al., 2009, p. 481).

Using an advanced eye-tracking paradigm, Emond and Shi (2024) evaluated 30-month-old French-learning toddlers' real-time interpretation of pronouns and reflexives. Toddlers demonstrated instant, accurate recognition that reflexives require local binding while pronouns require local freedom, adhering strictly to Principles A and B. Emond and Shi concluded: "Judgment tasks and act-out games can be challenging, as children's responses are sometimes affected by factors such as stimulus context, their understanding of the task, the way they prefer to play the game, etc." (Emond & Shi, 2024, p. 24).

Sanoudaki and Varlokosta (2015) evaluated Greek strong pronouns and clitics across two task formats: a TVJT and a Picture Selection (PS) task. Greek children exhibited adult-like mastery of Principle B without a PIP across both pronominal types. However, Sanoudaki and Varlokosta discovered that PS tasks systematically induced higher anaphoric coreference choices in both adults and children, warning that task mechanics directly shape observed response distributions.

Finally, Pinto and Zuckerman (2019) introduced the Coloring Book Task (CBT) as a method used to test children's language comprehension while mitigating known task-related biases. In the CBT, children color black-and-white sketched characters based on auditory prompts. Testing Dutch-speaking children, Pinto and Zuckerman found near-ceiling performance on both Principle A and Principle B (exceeding 90% accuracy for both), with no significant performance gap between reflexives and pronouns. The authors concluded that the CBT provides a clean window into child syntactic competence, reinforcing that the previous PIP results may be an artifact of heavy experimental task demands. Nevertheless, the status of the PIP remains an open empirical

question, as Kim (2024) notes that the phenomenon is "observed in different studies using diverse measures varying from offline picture judgment tasks to online listening tasks" (p. 462).

7. Conclusion and Future Directions

This comprehensive review highlights the complex and evolving landscape of generative language acquisition research. The historical view that children suffer from a prolonged syntactic delay in acquiring Principle B (the DPBE/PIP) has been repeatedly supported in some studies and refuted in others, generating a wide array of competing theoretical explanations ranging from pragmatic deficits to processing limits. Crucially, these persistent variations across the literature strongly suggest that task-induced artifacts play a genuine role in shaping children's performance. When experimental tasks control for computational complexity, eliminate forced binary choices, and satisfy pragmatic felicity conditions, young children across languages frequently demonstrate robust, adult-like mastery of pronominal binding constraints. Consequently, future research on the acquisition of Binding Theory, particularly on under-researched language varieties, must move away from traditional, highly demanding tasks (such as standard TVJTs or PS tasks) in favor of low-demand paradigms like the CBT and eye-tracking protocols. Within this context, the Arabic linguistic landscape represents a crucial but severely under-researched domain. Given the traditional methodologies used in the few existing studies on Arabic L1A, it is vital that future investigations of Arabic-speaking children adopt more controlled, low-demand paradigms. Ultimately, the story of Principle B acquisition lacks a single, definitive conclusion; even within individual languages, empirical findings remain strikingly mixed, underscoring that the PIP remains an open field of inquiry requiring continued research to uncover the underlying truth across these studies.

Funding: This research received no external.
Conflicts of Interest: The author declares no conflict of interest.
ORCID iD: https://orcid.org/0009-0004-8783-2201
Publisher's Note: All claims expressed in this article are solely those of the author and do not necessarily represent those of the author's affiliated organization, or those of the publisher, the editors and the reviewers.

References

[1] Al Kafri, A. (2019). *Similarities and differences between child and adult L2 learners in the acquisition of English reflexive and pronominal binding* [Master's thesis, University of Essex].
[2] Bolotin, N. (2002). Acquisition of binding in L1 Arabic. In S. R. Choudhury (Ed.), *L1 acquisition of Arabic* (pp. 211–218). John Benjamins.
[3] Chien, Y.-C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1*(3), 225–295.
[4] Chomsky, N. (1981). *Lectures on government and binding*. Foris Publications.
[5] Conroy, A., Takahashi, E., Lidz, J., & Phillips, C. (2009). Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry, 40*(3), 446–486.
[6] Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. MIT Press.
[7] Elbourne, P. (2005). On the acquisition of Principle B. *Linguistic Inquiry, 36*(3), 333–365.
[8] Emond, A., & Shi, R. (2024). The knowledge of binding principles in early child grammar: Experimental evidence from 30-month-old French-learning toddlers. *Language Acquisition, 31*(1), 1–28.
[9] Grodzinsky, J., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry, 24*(1), 69–101.
[10] Grolla, E. (2005). *Pronouns as elsewhere elements: Implications for language acquisition* [Doctoral dissertation, University of Connecticut].
[11] Guasti, M. T. (2002). *Language acquisition: The growth of grammar*. MIT Press.
[12] Jakubowicz, C. (1984). On markedness and binding principles. In C. Jones & P. Sells (Eds.), *Proceedings of the 14th annual meeting of the North East Linguistic Society* (pp. 154–182). GLSA.
[13] Kiguchi, H., & Thornton, R. (2004). Binding principles and ACD constructions in child grammars. *Cognition, 93*(3), 243–270.
[14] Kim, E., Montrul, S., & Yoon, J. (2015). The on-line processing of binding principles in second language acquisition: Evidence from eye tracking. *Applied Psycholinguistics, 36*(6), 1335–1374.
[15] Kim, J. (2024). Binding and co-reference: Pronouns in acquisition. *Journal of Child Language, 51*(2), 460–482.
[16] Lee, D., & Schachter, J. (1997). Sensitive period effects in Binding Theory. *Language Acquisition, 6*(4), 333–362.
[17] Matsuo, A. (1999). *Reciprocity and binding in early child grammar* [Doctoral dissertation, University of Connecticut].
[18] McDaniel, D., Cairns, H. S., & Hsu, J. R. (1990). Binding principles in the grammars of young children. *Language Acquisition, 1*(2), 121–139.
[19] McKee, C. (1992). A comparison of pronouns and anaphors in Italian and English acquisition. *Language Acquisition, 2*(1), 21–54.
[20] Mustafawi, E., & Mahfoudhi, A. (2005). The development of binding principles in Qatari Arabic. *Al-'Arabiyya, 38*, 19–43.

[21] Pinto, M., & Zuckerman, S. (2019). Coloring book: A new method for testing language comprehension. *Behavior Research Methods, 51*(6), 2609–2628.
[22] Reinhart, T. (1976). *The syntactic domain of anaphora* [Doctoral dissertation, Massachusetts Institute of Technology]. DSpace@MIT.
[23] Reinhart, T. (2006). *Interface strategies: Optimal and costly computations*. MIT Press.
[24] Reinhart, T. (2011). Processing or pragmatics: Explaining the coreference delay. In E. Gibson & N. Pearlmutter (Eds.), *The processing and acquisition of reference* (pp. 231–272). MIT Press.
[25] Ruigendijk, E., Baauw, S., Zuckerman, S., Polišenská, D., & Avrutin, S. (2010). Symmetry in comprehension and production of pronouns: A comparison of German and Hebrew. *Lingua, 120*(8), 1991–2005.
[26] Sanoudaki, E., & Varlokosta, S. (2015). Task effects in the interpretation of pronouns. *Language Acquisition, 22*(1), 40–68.Slabakova, R., Kempchinsky, P., & Rothman, J. (2017). Pronoun interpretation in the second language: Effects of computational complexity. *Studies in Second Language Acquisition, 39*(1), 1–31.
[27] Su, Y. (2020). Backward and forward anaphora in child and adult Mandarin Chinese. *First Language, 40*(4), 412–435.
[28] Varela, A. (1989). *The acquisition of binding domains in Spanish and English* [Unpublished manuscript]. Department of Linguistics, University of Connecticut.
[29] White, L. (1998). Second language acquisition and Binding Principle B: Child/adult differences. *Second Language Research, 14*(4), 425–439.