

# The Pronoun Interpretation Problem in child Hungarian

**György Rákosi**

Department of English Linguistics,  
University of Debrecen  
rakosi.gyorgy@arts.unideb.hu

**Enikő Tóth**

Department of English Linguistics,  
University of Debrecen  
toth.eniko@arts.unideb.hu

**Abstract:** In a number of languages, children have problems with the interpretation of pronouns if a potential local antecedent is present. There is an intensive debate on whether this effect is due to a delayed acquisition of Principle B, or it is the result of pragmatic or processing difficulties that children face in interpretation tasks. We conducted two experiments involving a picture-sentence verification task to investigate whether the Pronoun Interpretation Problem exists in Hungarian child language. We found that the Problem is present if the test sentences are given in isolation, but it disappears if a minimally coherent discourse is created. We argue that our results support the view that the binding principles are innate and do not need to be acquired, but children have problems with computing coreference options in certain contexts (Reinhart 2004; 2006; 2011). Coherent discourse allows children to accommodate pronouns with close to adult-like success because in this case they do not calculate local coreference possibilities for pronouns.

**Keywords:** Hungarian; child language; experiment; reflexive anaphor; personal pronoun

## 1. Introduction

There exists a well-known asymmetry in the acquisition of personal pronouns and reflexive anaphors in English child language. The problem has been first identified as such during the 1980s in the acquisition literature by Jakubowicz (1984); Wexler & Chien (1985) and Crain & McKee (1985), with the classic reference being Chien & Wexler (1990). While children more or less master the use of reflexive anaphors as reflexivity-markers at a relatively early age (by the age of 3–4), personal pronouns with potential local antecedents pose a substantial challenge for them. Thus while they perform well on the reflexive structure in (1a), they typically guess when

they are asked to judge the truth value of the pronominal construction in (1b) *qua* a description of a single-participant reflexive event.

- (1) a. Papa Smurf washed himself.  
 b. Papa Smurf washed him.

In other words, while Principle B of the Binding Theory prohibits the co-construal of *him* and *Papa Smurf* in (1b) for adults, children can accept this interpretation in an experimental setting. The problem persists until the early school age (7–8). This phenomenon was termed the **Delay of Principle B Effect**, or DPBE for short.

Subsequent research has discovered two important facts concerning this acquisitional phenomenon. First, it is not universal. While the DPBE has been reported in English, Icelandic, Russian or Dutch, experimental evidence is available that it is absent in French, Italian, Spanish, Catalan, Norwegian and German (see Rooryck & Vanden Wyngaerd 2015 for a recent overview). Second, the problem only manifests itself in full in comprehension, but not in production (see Bloom et al. 1994; de Villiers et al. 2006; Spenader et al. 2009). Accordingly, the earlier term **the Delay of Principle B Effect** has been replaced by the term **Pronoun Interpretation Problem** (PIP) in some of the pertinent literature (see Hamann 2011 for a discussion). We adopt this latter term here.

In this paper we report the results of our two experiments that we conducted to investigate whether the Pronoun Interpretation Problem is present in Hungarian child language or not. The children were asked to provide truth value judgements of transitive reflexive and pronoun constructions in a sentence–picture verification task in both experiments. There was only a minimal difference in design between the two experiments: the test sentences were presented in isolation in the first, whereas in the second they were preceded by a short lead-in sentence containing a topic that could serve as a natural discourse antecedent for the pronoun in the test sentence. We adopted the second design from Spenader et al.’s (2009) study of the Problem in Dutch child language. They found that if a coherent discourse is created, then children perform much better on Principle B tasks and the Pronoun Interpretation Problem essentially disappears. Our results repeat their findings inasmuch as our first experiment established the presence of the PIP in Hungarian, but the design of our second experiment brought about a radical improvement in performance in the pronoun condition.

There is an extensive literature on the nature and the empirical scope of the Pronoun Interpretation Problem, and the debate on the adequate

description and analysis of the data has recently received new impetus from studies that reinterpret the earlier findings. Our results address some of the fundamental questions that have shaped this debate.<sup>1</sup> We concentrate on three of these questions, which are directly relevant to our findings. First, the methodology used in many of the experiments studying the PIP has been criticised by Elbourne (2005) and Conroy et al. (2009). In fact, Conroy et al. (2009) come to the conclusion that if certain methodological flaws receive proper attention and the experiments are repeated with an improved design, then the Problem disappears in its classic form and only a *residue* remains (with 70%–90% adult-like performance on the coreference option in (1b)). In our first experiment, children did not appear to perform much better than the classic chance-level pattern that Chien & Wexler (1990) found, while they were close to adult-like performance in the second experiment. We argue that this discrepancy is not experimental noise, but it has its own theoretical significance. This leads to two further issues that we focus on here. One concerns the innateness of the rules of binding and coreference that children may or may not utilize in these tasks. It has been claimed recently that Principle B may have to be acquired and thus its acquisition is delayed in the true sense of the word in languages that have the PIP (see Elbourne 2005 and Rooryck & Vanden Wyngaerd 2015). The results of our two experiments, taken together, support an analysis which assumes that Principle B is innate (see, among others, Grimshaw & Rosen 1990; Chien & Wexler 1990; Thornton & Wexler 1999; Grodzinsky & Reinhart 1993; Reinhart 2004; 2006; 2011). What needs an explanation then is why the PIP surfaced in the first experiment, which is the third issue that we address in this paper. We adopt Reinhart's (2004; 2006; 2011) account, which rests on the observation that children have problems only with **coreference**, but not with **variable binding**. The PIP is the result of the fact that when children try to compute coreference options in the classic pronoun-false condition, they fail due to processing limitations. In other words, they know the rules but they lack the brain capacity to evaluate alternative derivations on the fly to establish whether coreference is licit in the particular case of (1b) or not. As a result, they start guessing. We argue that the results of our second experiment can be interpreted to show that coherent discourse allows children to opt out of this computation in the presence of a salient discourse antecedent and their performance becomes almost adult-like.

<sup>1</sup> We refer the reader to Hamann (2011) for an excellent overview of the history of this debate and of the state-of-the-art.

The structure of the paper is as follows. In section 2, we provide a brief overview of recent advances in research on the empirical and theoretical aspects of the Pronoun Interpretation Problem, focusing on Reinhart's (2004; 2006; 2011) account. This discussion serves as a background to the presentation of the key design features of our experiments. We also give a short summary of the pertaining literature on Hungarian. Sections 3 and 4 respectively present the results of our first and second experiment, as well as the pertinent discussion of the findings. We give a summary of our conclusions in section 5.

## **2. The empirical and theoretical background to the two experiments**

### **2.1. Overview**

In this section, we first describe and discuss the processing account of the PIP as established by Grodzinsky & Reinhart (1993) and later developed further by Reinhart (2004; 2006; 2011). Then we overview some recent critiques of this account. While these raise important concerns, we argue that they do not refute the Reinhart–Grodzinsky account, which we adopt as our analytical frame for our experiments. We also use this discussion to set up the empirical background for our investigation of the PIP in Hungarian child language.

### **2.2. The processing account of Reinhart (2004; 2006; 2011)**

The most influential account of the Pronoun Interpretation Problem (or the Delay of Principle B Effect) was presented by Grodzinsky and Reinhart (1993) and later updated in Reinhart (2004; 2006; 2011). We give a summary of the relevant aspects of this account on the basis of Reinhart (2006), which is a monograph-length treatise of the larger grammatical context for a proper evaluation of phenomena that involve the PIP itself.

The starting point of the argument is the recognition that children generally do well on some tasks of anaphora resolution, whereas they fail on others. Children typically perform well on anaphoric dependencies that involve true binding. They know that a reflexive needs a local antecedent, so they can accept (2a) only as a description of a reflexive event. Chien and Wexler (1990) showed that children also do well with quantified antecedents: they know that *him* cannot be bound in (2b), so they dominantly reject (2b) on the reflexive reading. But they are ready to accept (2b) if

the pronoun has a deictic antecedent outside of the local domain. Children also do well on referential antecedents as long as the construction is a description of a two-participant event (2c).

- (2) a. Papa Smurf washed himself.  
 b. Every Smurf washed him.  
 c. Papa Smurf washed him. (Papa Smurf  $\neq$  him)

They form these judgements because they know Principle A and Principle B, which are inborn.<sup>2</sup>

The PIP surfaces if the pronoun construction is meant to be a description of a reflexive event. In this case, children typically show a guessing pattern.

- (3) \*Papa Smurf washed him. (Papa Smurf = him)

Under special circumstances, pronouns can have local antecedents via the mechanism of (accidental) coreference in adult English. But this option has to be constrained to restrict such construals to specific contexts only, and to preserve the original coverage of Principle B. Reinhart (1983; 2006) argues that this coreference option is only available if the resulting interpretation is truth-conditionally distinct from what could be obtained through binding. This is the case in (4): (4a) is true if no other smurfs washed themselves (i.e., (4a) describes a self-washing-event), whereas (4b) is true if no other smurf washed Papa Smurf (i.e., (4b) describes an event in which two distinct roles are accidentally born by the same individual in the same event).

- (4) a. Only Papa Smurf washed himself.  
 b. Only Papa Smurf washed him. (Papa Smurf = him)

<sup>2</sup> A standard formulation of the binding principles suffices for the purposes of this paper (see Chomsky 1981). Principle A requires anaphors to be bound within their local domain, whereas Principle B prohibits pronouns from being locally bound. While many alternative proposals have been made since Chomsky (1981) (including Reinhart & Reuland 1993), it is standard in the acquisition literature to use the terms **Principle A** and **Principle B** in a somewhat loose GB-theoretic sense. Since Reinhart (2006) is no exception to this terminological convenience, we also follow this practice and use these two terms in this sense.

What licenses coreference in (4b) and what rules coreference out elsewhere in the local domain is what Reinhart (2006, 210) calls Rule I.<sup>3</sup>

(5) **Covaluation Rule I**

$\alpha$  and  $\beta$  cannot be covalued if

- a.  $\alpha$  is in a configuration to bind  $\beta$ , (namely,  $\alpha$  c-commands  $\beta$ ) and
- b.  $\alpha$  cannot bind  $\beta$  and
- c. the covaluation interpretation is indistinguishable from what would be obtained if  $\alpha$  binds  $\beta$ .

Let us now consider in detail how Rule I licenses the coreference interpretation in (4b). First we check (5a). The subject noun phrase *Papa Smurf* ( $\alpha$ ) is in a configuration to bind *him* ( $\beta$ ), so this condition is satisfied. At the same time, it also holds that *Papa Smurf* cannot bind *him*. This is so because *him* – unlike *himself* – is lexically specified to be a pronoun, and thus is subject to the prohibitive constraint of Principle B in the local context. Finally, (5c) is **not** satisfied in the case of (4b), because the coreference interpretation is truth-conditionally distinct from what we could obtain from an alternative binding derivation. Rule I therefore does not apply, and coreference is not blocked. Note that (5c) requires the comparison of two alternative derivations at the interface between the computational system and an interpretive module of grammar. Reinhart (2006) refers to this extra workload as **reference-set computation**, and argues that this procedure is also involved in the computation of quantifier scope, focus and implicatures.

Thus, Reinhart's analysis draws a crucial distinction between binding and coreference, as two distinct modes of establishing anaphoric dependencies. The former is governed by Binding Theory and is a matter internal to the computational system, and the latter is licensed by Rule I, which is operative at the interface. Rule I, in essence, disallows coreference in case it would produce the same semantic result as binding. The underlying assumption is that if one interpretation is ruled out by the computational system, then we do not want to reintroduce the self-same interpretation by another mechanism at the interface. There has to be an interpretation that is not available otherwise for coreference to be licensed in constructions that fall within the scope of Rule I. The transitive construction we investigate here is one such construction.

<sup>3</sup> Reinhart (2006) replaces the term *coreference* with *covaluation* for principled reasons, which are not directly relevant for us. We continue using the former term in this paper.

So, for a locally coreferential pronoun to be licensed, an alternative binding derivation must be constructed to check whether the target pronoun construction is indeed truth-conditionally distinct. Just like the binding principles, Rule I is inborn, so children know what they are supposed to do. But this task involves the simultaneous processing of two derivations, which, Grodzinsky & Reinhart (1993) and Reinhart (2006) argue, children are not capable of due to their working memory limitations. Thus they typically start guessing if they are to interpret (3) as a description of a single-participant event, and the Pronoun Interpretation Problem appears. It is important to note that this account requires the comparison of two hypothetical derivations and does not say much about how these two derivations could be realised morphosyntactically in a given language. In other words, it compares coreference and binding **interpretations**, and not two types of **pronominals** (or at least not in any direct manner). In this sense, this account predicts the universality of the PIP in the absence of independent constraints to the contrary. All that is required is a pronoun that can be referential. *Him* is such a pronoun, whereas the reflexive *himself* is lexically specified to be an anaphor and is thus not subject to Rule I.

### 2.3. Pragmatic considerations and their relevance for our experiments

Chien and Wexler (1990) present a pragmatic analysis of the PIP, and Thornton and Wexler (1999) further develop this account. They argue that children know Principle B, but there is available input in the adult language for English children to believe that pronouns can take local antecedents in certain cases. (6) is from Chien & Wexler (1990, 256):

(6) That must be John. At least he looks like him. (John = he = him)

This is an example that lies within the coverage of Reinhart's Rule I. Chien and Wexler (1990), as well as Thornton and Wexler (1999), argue, however, that what licenses examples of this sort is a pragmatic rule. In essence, such pronouns are licit if they somehow present an aspect or a **guise** of their referent that is distinct from that of the antecedent (see Heim 1998). In (6), the speaker identifies an individual on the scene and compares it to the person he otherwise knows to be John. On the Thornton & Wexler (1999) account, children have not yet mastered how to handle such cases and extend the use of the pragmatic conditions that allow different guises for the same individual. This results in the Pronoun Interpretation Problem.

Reinhart (2006; 2011) points out that this account does not predict why children perform at around the chance level in Principle B environments. In fact, if they can freely extend the base of the adult pragmatic rule that licenses (6), we may expect children to accept locally coreferential pronouns well above the chance level. But this does not happen in the classic experimental setup.

We may add to this that this pragmatic account requires the presence of adult input which triggers the building up of a pragmatic rule in the acquisition process. Hungarian adults, however, strongly disallow locally coreferential pronouns. So adults find (7) unacceptable in any context if the pronoun and the subject are meant to corefer.

- (7) \*Csak János lát-t-a ő-t. (János = őt)  
 only John.NOM see-PAST-3SG.DEF-OBJ he-ACC  
 ‘Only John saw him.’

This is a somewhat curious fact, given that the Hungarian object pronouns are not clitics and they can receive focus stress.<sup>4</sup> But even focussing does not save (7). It is noteworthy in this context that besides the default reflexive anaphor, which is morphologically complex, Hungarian has even more complex reflexive anaphors and these do license local coreference (see Rákosi 2011). Thus Hungarian children may receive evidence for the existence of local coreference within the language, but, crucially, this does not come from object pronouns. If they still show the PIP, then it is a problem for the Thornton & Wexler (1999) account since there is no direct trigger in the input for the construction of the sort of pragmatic rule that English children are supposed to overgeneralize in the case of pronouns. In contrast, Reinhart’s Rule I is inborn and is therefore universal. So if Hungarian object pronouns are referential – and enough evidence is available for children that they are –, then children may accept them in the transitive construction represented by (7), despite the lack of direct evidence in adult language.

The Optimality-Theoretic account developed by Hendriks and Spender (2005/2006) and Spender et al. (2009) is another analysis that claims that pragmatic factors may directly intervene – albeit in the direction of eliminating the PIP altogether. They build on the idea that pro-

<sup>4</sup> Strong stress on the pronoun facilitates coreferential readings in English, even if it is not a necessary condition for coreference (see Reinhart 2006; 2011). We return to clitic pronouns in 2.5 in the general context of the PIP and in the specific context of Hungarian.



nouns compete with reflexives: other things being equal, reflexives are the preferred markers of anaphoric dependencies, and when reflexive anaphors fail, it is then that we can use a pronoun.<sup>5</sup> When adults interpret a clause that contains an object pronoun, they establish the disjoint interpretation via the reasoning that Spenader et al. (2009) characterise as follows: “Adults hear a pronoun, consider that the speaker could have produced a reflexive and recognize that a reflexive is unambiguously co-referential with the subject. They realize that if the speaker intended a co-referential interpretation, the speaker would have used a reflexive” (*op.cit.*, 32). Adults are capable of what they analyse as **bidirectional optimization** in their OT model, which essentially means that they can take into account the opposite perspective in communication. Children can use the grammar in one direction only, which is modelled as unidirectional optimization from meaning to form (production) or from form to meaning (comprehension). So children cannot reconstruct the motivation why their interlocutor used an object pronoun instead of a reflexive in a transitive construction like (3), and, consequently, they can accept pronouns as descriptions of reflexive situations.

Spenader et al. (2009) claim that their model makes the same predictions for pronouns as Reinhart’s (2006) analysis, except for the influence of one important pragmatic factor. When a favourable discourse context is available that facilitates the choice of an extra-sentential antecedent, then children are expected to improve and will not show the classic PIP-pattern. They compare what they call the **Classic Condition** of Chien & Wexler (1990) and their **Single Topic Condition** (8a vs 8b):

- (8) a. This is Goldilocks. This is Mamma Bear. Is Mamma Bear washing her?  
 b. This is Goldilocks. Is Mamma Bear washing her?

The reason why many children may opt for the co-construal of *Mamma Bear* and *her* in (8a) is that Mamma Bear is the closest topic and thus the closest possible antecedent for the pronoun. If only a single topic is included in the lead-in and it serves as an extra-sentential antecedent for the object pronoun, children should perform better (8b). And so did the Dutch children Spenader et al. (2009) investigated in their experiment. While the children’s correct interpretation of pronouns was only at 69% in the Classic Condition, they improved to 83% in the Single Topic Condition

<sup>5</sup> See Safir (2004) and Rooryck & Vanden Wyngaerd (2011) for two different proposals that also rely on the idea that competition is the driving force behind the observed distribution of pronominals.

(*op.cit.*, 41). This leads them to conclude that children are highly proficient at using pragmatic clues in anaphora resolution tasks, and coherent discourse may actually eliminate the Pronoun Interpretation Problem in its classic form.

The ability to use pragmatic clues should be universally available to children. For this reason, we decided to adopt this key design feature of Spenser et al. (2009). In our first experiment, we provided no immediate linguistic context for the test sentences, and we provided a single topic lead-in in the second experiment. We expected children to show improved performance if such minimal pragmatic clues are provided. This is exactly what happened: children performed significantly better in the second experiment.

Spenser et al. (2009, 50) point out that it is unclear how Reinhart's (2006) account can be extended to capture the ameliorating effect of coherent discourse. We suggest below that children possibly do not run Rule I in such contexts. They opt out from performing the costly reference-set computation, and abide by Principle B and the favourable discourse structure in evaluating object pronouns.

## 2.4. Methodological issues

Several authors have raised their concerns about the methodology of earlier experiments and questioned the validity of the conclusions reached in previous studies. As an illustration, consider the case of what Elbourne (2005) refers to as the **Quantificational Asymmetry**. First discussed in detail in Chien & Wexler (1990), the asymmetry arises between quantificational and referential antecedent constructions, as in the following example ((2b) and (2c) repeated):

- (9) a. Papa Smurf washed him.  
       b. Every smurf washed him.

As discussed above, Chien and Wexler (1990) found that children typically guess when they face a task in which they have to apply the pronominal description in (9a) to a single-participant reflexive situation, i.e., their performance is around 50%. They, however, perform much better if the antecedent is quantificational (9b), since in this case they tend to reject the bound interpretation of *him*. In other words, there is no Pronoun Interpretation Problem with quantificational antecedents. The Quantificational

Asymmetry was an important motivation for the analysis that Grodzinsky & Reinhart (1993) and Reinhart (2004; 2006; 2011) propose (see 2.2).

Elbourne (2005), however, argues that the Quantificational Asymmetry is most probably a side effect of many existing experimental designs. Children rejected the bound variable interpretation of the pronoun in the quantificational construction (9b) in a number of experiments simply because another, more salient deictic antecedent was available in the narrative or in the picture that accompanied the test sentences. He refers to this mode of explanation as the **Salience Hypothesis**, and similar concerns are raised in Conroy et al. (2009) and Spender et al. (2009). Suppose, for example, that (9b) is presented after a short story that involves the smurfs and Gargamel. Gender issues aside, the children picked Gargamel as the referent of *him* in (9b) because Gargamel played a more salient role in the story, and thus they rejected (9b) as the description of a reflexive situation not because they cannot accept the bound variable reading of *him*, but simply because the smurfs were less prominent in the story. Similar considerations apply to picture verification tasks in which a potential deictic antecedent was depicted in a more prominent manner than the quantificational antecedent. The Quantificational Asymmetry thus may only be an illusion, and if it is, argues Elbourne (2005), then we have no firm evidence that children know Principle B. In other words, the Pronoun Interpretation Problem is the result of the lack of Principle B.

While these methodological concerns are important, the discussion over the validity of the empirical data is far from being over. De Villiers et al. (2006), for example, did find evidence for the Quantificational Asymmetry in their experiment. In any case, Conroy et al. (2009) point out that if there indeed is no asymmetry, that removes one argument from those supporting the processing account of Reinhart (2004; 2006; 2011), but it does not refute it per se.

We decided not to include the quantificational condition in our experiments, and used only proper names or definite descriptions in the non-pronominal argument slots in the sentences. The logic behind this decision was that if the PIP is present in Hungarian, then it is the referential antecedent condition anyway that will provide relevant – and valid – data. Since our immediate goal is to investigate the PIP, we leave it for another occasion to test whether the Quantificational Asymmetry is present in Hungarian child language.

We note here finally that Conroy et al.'s (2009) thorough criticism of earlier experiments leads them to reject the existence of the Pronoun Interpretation Problem altogether. But even they acknowledge that what

they call a **residue** remains in the range of 10% and 30% incorrect answers in the pronoun condition. Their conclusion raises a number of questions, however, as Hamann (2011) rightly observes. Why is it, for example, that adult controls do not have the same problems as children do even if no elaborate context is provided in the experiment? Or why is it that children perform better in the reflexive conditions, again, irrespective of the experimental design? These are relevant concerns in the contexts of our experimental results, too, and they strongly suggest that the PIP is not an experimental artefact. We note furthermore that we found that a relatively small change in the design brought about radical improvement in performance in the pronoun conditions in our two experiments. The difference between the results of the two experiments is not simply an unwelcome side-effect of experimental methodology, but an important fact to be accounted for. We discuss this issue in more detail in section 4 after the presentation of the results of the two experiments.

## 2.5. A note on cross-linguistic variation

We noted in the introduction that there is an interesting cross-linguistic variation in the appearance of the PIP. Whereas it has been reported in English, Icelandic, Russian or Dutch; experimental evidence is available that it is absent in French, Italian, Spanish, Catalan, Norwegian and German. It has been argued that the reason why no PIP-effect has been found in Romance is that the Romance clitics are structurally and referentially deficient (see, for example, Cardinaletti & Starke 1995 for one such approach, as well as Hamann 2011 for an overview of the pertinent literature). If they are lexically deficient to license coreferential readings (at least in the usual case), then Reinhart's (2006) Rule I (5) cannot apply and no PIP arises. As Rooryck and Vanden Wyngaerd (2015) point out, this does not explain why the PIP has been found to be absent in Norwegian and German, too, since these languages do not have clitics in the Romance sense. We cannot discuss this issue here in more detail, but make the following remarks in the context of our Hungarian experiments.

In a recent experiment, Hartman et al. (2012) have found a significant difference between the interpretation of full and reduced pronouns in child English. Consider the following minimal pair:

- (10) a. Papa Smurf washed him.  
 b. Papa Smurf washed 'm.

Children performed much better in the reduced pronoun condition (10b) than in the full pronoun condition (10a). Whereas they showed the classic chance pattern in the latter (correct responses: 52.8%), their performance was much better in the former (correct responses: 80.6%). This suggests that the presence of the PIP is not conditioned primarily by language type, but by the type of the pronoun employed in the test sentences.

The Hungarian object pronouns and reflexive anaphors are not like the Romance clitics. They can be focussed, conjoined and they can freely occupy different positions in the clause structure. They can also appear in isolation. However, Hungarian is known to be a language that allows for object *pro*-drop (see É. Kiss 2002). Object pronouns (and reflexives) can be omitted if they play no designated discourse function, i.e., when they are not the topic or the focus of the clause. But *pro*-drop is not necessary even in the absence of such a function and discourse-anaphoric pronouns may appear in the post-verbal domain with no observable truth-conditional differences with respect to the *pro*-dropped variant. Consider the following two sentences for illustration:

- (11) a. Ott van János. Te is lát-od?  
 there is John.NOM you.NOM also see-2SG.DEF-OBJ  
 ‘John is over there. Do you also see him?’
- b. Ott van János. Te is lát-od ő-t?  
 there is John.NOM you.NOM also see-2SG.DEF-OBJ he-ACC  
 ‘John is over there. Do you also see him?’

*Pro*-drop (11a) or the inclusion of the object pronoun (11b) are both possible in this context. The Hungarian minimal pair thus resembles the English one in (10), but this time it is the difference between a *pro*-dropped and a full pronominal form that is important. Other things being equal, and provided the PIP does surface in the full pronoun condition in Hungarian, we expect Hungarian children to perform better in the *pro*-drop condition and possibly reject such sentences as descriptions of reflexive events. We did not, however, include this condition in the current experiments. The logic is the same as above: if the PIP exists in Hungarian, we expect it to be present when full pronouns are involved. Since our fundamental question in this paper is whether Hungarian children show the PIP or not, we only used full pronouns (and full reflexives) in our experiments. We intend to investigate children’s performance in the *pro*-drop condition in a future experiment.

## 2.6. The literature on Hungarian

We are aware of one study that addresses Hungarian data that fall within the scope of the PIP.<sup>6</sup> Czingráber (1999) investigated how children and aphasics interpret pronouns and reflexives in a selection of syntactic constructions.<sup>7</sup> One of the structures she includes is the transitive construction with a potential subject antecedent and a reflexive or pronoun object. Data were collected from 16 children in a picture verification experiment, 8 from two kindergarten age groups and 8 from the first two classes of primary school. The primary school children performed in an adult-like manner in the pronoun condition, whereas the kindergarten children showed variation between 3 and 9 correct answers out of 10. Each group did well in the reflexive condition. The children watched a picture and had to evaluate the truth of the accompanying sentence that was presented in isolation, without a lead-in.

Czingráber's (1999) results suggest that Hungarian children show the Pronoun Interpretation Problem at least till around 6 years of age. However, the number of kindergarten subjects involved in this experiment was relatively low (8 children) to be able to draw firm conclusions. Our aim with the two experiments that we report here was to include a greater number of subjects and two alternative designs to gain deeper empirical substantiation for the analysis and a better understanding of what Hungarian children do when they try to interpret these sentences.<sup>8</sup>

## 3. Experiment 1: testing sentences without a lead-in

### 3.1. Hypotheses

As discussed earlier, our primary concern was whether our experimental data provide evidence for the Pronoun Interpretation Problem in child Hungarian or not. More specifically, first we wanted to see whether the

<sup>6</sup> Pléh (1998) investigates the processing of intersentential anaphora in contexts where no potential local antecedent is available. See Bencze (2014) for an overview of adult and child language studies of intersentential anaphora in Hungarian.

<sup>7</sup> Czingráber (1999) is the university thesis of Márta Czingráber, written under the supervision of Zoltán Bánréti.

<sup>8</sup> The aphasic patients involved in the Czingráber (1999) study (three Broca's aphasics and one Anomic aphasic patient) and the two Broca patients in Bánréti (2006) study showed variation between 50-90% of correct answers. This may be interpreted as further evidence that Hungarian has the PIP, but we leave this issue open here.

classic results supporting the PIP can be reproduced for Hungarian. Hence, our aim was to find empirical evidence in favour of or against the presence of the PIP in Hungarian and to explore possible differences with respect to age. We tested children aged 4–5 and 6–7 years, respectively. Based on the theoretical assumptions presented above, the following hypotheses have been tested.

**Hypothesis 1** The PIP will be detected in the absence of facilitating discourse, i.e., both age groups involved in the experiment will perform better when interpreting reflexives.

**Hypothesis 2** Children from the preschooler group (aged 6–7) will outperform younger children from the intermediate group (aged 4–5) in each condition, i.e., older children will show a more adult-like pattern.

### 3.2. Participants

40 children participated in the experiment from the intermediate and the pre-schooler group of a local kindergarten.<sup>9</sup> The children received a small toy at the end of the session as a reward for their participation in the experiment. We tested 20 children from both groups. Six children were excluded from the analysis, because they made three or more than three mistakes in the case of the six control sentences. These children either consistently said *yes* or *no* through the experiment, or they appeared to be guessing in each condition, which suggests that they did not understand the task. Four children were discarded from the intermediate group, and two from the preschooler group. Table 1 summarizes the gender and age data of the remaining children in both age groups.

**Table 1:** The participants of Experiment 1

	Boys	Girls	Total	Mean age	Age range
Intermediate	10	6	16	4;5	4;2–4;11
Preschooler	8	10	18	6;8	6;0–6;12

Most children in the preschooler group were spending an extra year at the kindergarten, partly in preparation for the primary school. This explains

<sup>9</sup> This kindergarten is in Debrecen, Hungary.

why the age difference between the two groups is roughly two years on the average. We also had an adult control group consisting of nine (four men and five women) first year university students, with no background in linguistics and with no previous understanding of the phenomena that we investigated.

### 3.3. Materials

We employed a sentence-picture verification task. The test sentences were all transitive constructions. The object in the target sentences was either the 3SG accusative pronoun *őt* ‘him’ (12a), or the 3SG accusative form of the reflexive anaphor (12b).<sup>10</sup>

- (12) a. Nyuszi        le-locsol-ja                    ő-t.  
 Rabbit.NOM PRT-sprinkle-3SG.DEF-OBJ he-ACC  
 ‘Rabbit is sprinkling him.’
- b. Nyuszi        le-locsol-ja                    magá-t.  
 Rabbit.NOM PRT-sprinkle-3SG.DEF-OBJ himself-ACC  
 ‘Rabbit is sprinkling himself.’

The subject was a name or a definite description in each sentence. We tested four different types of sentence-picture combinations: pronoun-true, pronoun-false, reflexive-true and reflexive-false. The experiment included 6 sentences with pronoun objects, each used once for a description of an other-directed event (true) and once for a self-directed event (false). Similarly, there were 6 sentences with reflexive objects, used once to describe a self-directed event (true) and once for an other-directed event (false). We had therefore 4 experimental conditions equally spread across 24 target sentences. The four conditions are illustrated in Figure 1. In addition, there were 6 control sentences which contained a non-pronominal definite noun phrase in both the subject and the object position. Thus the test included 30 sentences altogether.

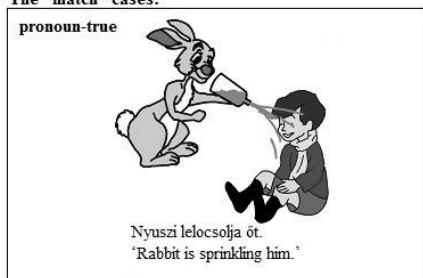
The sentences contained either of the following 6 causative verbs: *lelocsol* ‘sprinkles’, *megkötöz* ‘ties up’, *fejbe vág* ‘knocks on the head’, *megéget* ‘burns’, *megszúr* ‘stings’, *besároz* ‘makes muddy’. The characters in the

<sup>10</sup> Hungarian has no grammatical gender, so the gender feature was not relevant in our study. With the exception of Christopher Robin, Snow White and the Prince (and one of the Seven Dwarfs), each character was an anthropomorphic non-human character from a tale that the children are familiar with (e.g., Eeyore, or the wolf from the story of *Little Red Riding Hood*).

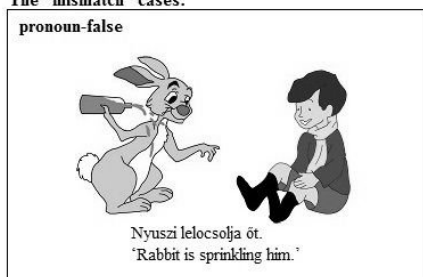


transitive subject and object argument position were drawn from the following stories: Winnie-the-Pooh, The Mole, Snow White, Donald Duck, Little Red Riding Hood and Süssü the Dragon (this latter is a Hungarian tv series that all children were familiar with). The pictures that we used contained two characters of equal graphical prominence to avoid the sort of biasing methodological flaws that Elbourne (2005) raised concerns about (see 2.4).<sup>11</sup> Each target picture depicted one of two types of situations. In one type, one character performs an action on the other. In the other type, one character performs a reflexive action on himself and the other character stays idle.

**The “match” cases:**



**The “mismatch” cases:**



**Figure 1:** The experimental conditions in Experiment 1

<sup>11</sup> We are grateful to Ágnes Lukács and Bence Kas for granting us access to their experimental material. We used some of the pictures they prepared for their anaphoric pronominal experiments with Hungarian SLI children. They adopted van der Lely and Stollwerck's design (1997), which we did not follow and changed the test sentences accordingly. At the time of the writing of this manuscript, they have not yet published their results (see Ladányi et al. in preparation).

### 3.4. Procedures

We used a within-subjects design. Participants were tested individually, they sat in front of a laptop with one of the experimenters who held a frog puppet and used it to communicate with them. We decided to use the puppet because it helps to create a friendly atmosphere and the children may feel more like being part of a game rather than a test (see Crain & Thornton 1998 for more on this).

Each individual session started with a quick introduction of two parts. First we watched pictures of each character that appears in the sentences to make sure that children know them. Second, we gave verbal instructions and did a short trial of the sentence-picture verification task to make sure the children know what they are supposed to do. The children were shown pictures and listened to accompanying sentences, and they were asked to make appropriate truth value judgements. If the children gave a *yes* answer, we asked them to give a (wooden) fruit to the frog, and if they gave a *no* answer, we asked them to indicate this by giving the frog a piece of cloth.

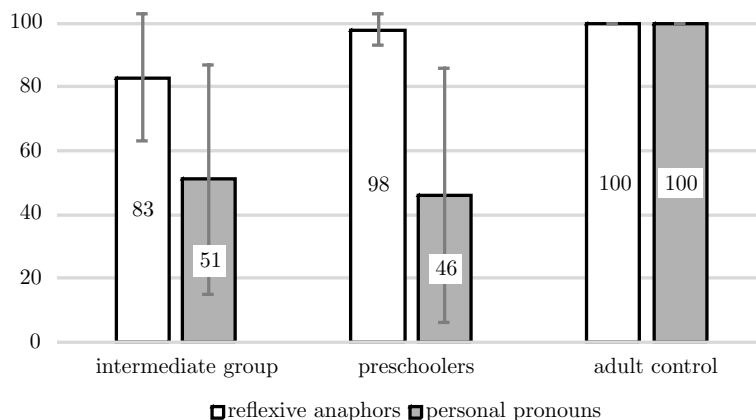
The pictures were shown on a laptop, which we also used to play the pre-recorded test sentences.<sup>12</sup> The children could hear the sentences 2 seconds after the pictures appeared on the screen. To run the experiment and to record the responses, we used the Pypres toolkit developed by Daniele Panizza. We told the children that the frog played with the computer and he jumbled up the pictures and the sentences so that now some of them do not match. We asked the child to help the frog put things back in order. We adopted this idea for a cover story from Spenader et al. (2009).

The introductory part included 6 sentences that contained no pronouns or reflexives. The 30 trials (24 target, 6 control) came after this introductory part. We randomized the order of the test items for each participant. Yes/no responses were registered for all test items on the laptop by the experimenter and they were also noted down on paper by the assistant. We spent 20–25 minutes with each child, and no session included more than 10 children.

<sup>12</sup> The linguistic material was pre-recorded by an assistant. The instruction was to use neutral intonation and to place no emphatic stress on the pronoun or the reflexive.

### 3.5. Results

Figure 2 shows the mean percentages of correct (adult-like) answers of reflexives and pronouns in the case of the two target groups and the adult control group.



**Figure 2:** Results of Experiment 1: mean correct interpretation of reflexives and pronouns

The adult control group did equally well in both conditions (100%). The results of the children are close to the adult pattern in the case of reflexives, but fall far from it in the case of pronouns. The preschoolers did better in the reflexive condition than children in the intermediate group, and this difference is significant (independent samples *t*-test (equal variances not assumed):  $t(16) = -2.89$ ,  $p < 0.05$ ). The difference between the pronoun and the reflexive conditions is also significant within both age groups (paired samples *t*-test comparing the number of correct answers, intermediate group:  $t(15) = 2.908$ ,  $p < 0.05$ , preschooler group:  $t(17) = 5.24$ ,  $p < 0.01$ ). The difference between the performance of the two age groups in the pronoun condition is not significant. These results show that children find it difficult to interpret pronouns, but they process reflexives well. This means that the PIP has been detected in Hungarian. Children's performance did not improve with age in the pronoun condition, and pronouns are apparently problematic for both age groups. It is noteworthy that individual variation as indicated by variance is quite remarkable in the pronoun condition in both kindergarten groups, while it is smaller in the reflexive condition (see Figure 2).

### 3.6. Discussion

The primary aim of the experiment was to test whether Hungarian children show the PIP, i.e., to test whether there is a pronoun/reflexive interpretation asymmetry. The results from the reflexive condition confirmed our expectations in the sense that we did not anticipate poor performance here. The preschoolers in fact performed at almost adult-like levels (98%), while children in the younger age group provided correct answers in 83% of the trials with a considerably larger individual variation. As mentioned above in 3.5, the difference between the results of the two age groups is significant. During the experimental trials, when children provided a wrong answer, we asked them to explain what they think was wrong. In these cases younger children typically reported referential issues. So, for example, they did not accept the description “The wolf is burning itself” for a reflexive action because they argued that “In fact it is burning its tail”. We suspect that these responses may be late reflexes of early non-adult-like uses of the reflexive. Probably these children were still too young to master the use of reflexives properly.<sup>13</sup>

However, the relatively good performance on reflexives is a noteworthy fact in the context of our investigation of the Pronoun Interpretation Problem. Since children also performed well when the reflexive description was false, it cannot be the case that they are generally perplexed upon being confronted with a picture and a description that do not match. Children rejected false reflexive descriptions without hesitation and in a definitive manner. This contrasts with how some of them reacted when they heard a sentence in the pronoun-false condition. Some children were overtly perplexed when they heard the construction *Rabbit is sprinkling him* applied to a reflexive event. This cannot be the result of a general tendency to fail on tasks that involve an ungrammatical interpretation, given that the children did much better in the reflexive-false condition.

The finer distribution of the correct answers is especially relevant now, and Table 2 contains the breakdown of the data in the pronoun-false condition. The adult-like behaviour is the rejection of the test sentences, since the picture depicted a reflexive event in these cases.

We compared the performance of the two groups again based on the frequency data presented in Table 2 using the chi-square test. In accor-

<sup>13</sup> Chien and Wexler (1990) also report a gradual improvement with age in some of the reflexive conditions they test in English. This does not, however, change the larger picture. Children are consistently better in the reflexive conditions than in the pronoun conditions across the experiments reported in the literature.

**Table 2:** The number of children giving *no* answers in the given range in Experiment 1, pronoun-false condition

Number of <i>no</i> answers in the pronoun-false condition	6–5 adult-like	4–2	1–0
Intermediate	14 (25%)	8 (50%)	4 (25%)
Preschooler	5 (28%)	4 (22%)	9 (50%)

dance with the results described above the test does not show a significant difference between the two age groups. However, there is a relatively high ratio of preschoolers who accepted the false descriptions as being true, which actually resulted in a group-level performance that is a little worse than that of the intermediate group (see Figure 2). This may be the result of a possibly conscious decision to play what they think is a safe strategy in a situation in which they are confused by the task. Other things being equal, it is safer to say *yes* than *no* to an adult.<sup>14</sup> So they may have said *yes* to please the adult experimenter, and it is possible that at least some older children have acquired this as a social strategy – something that is still absent in the younger age group.<sup>15</sup> It has to be emphasized here that individual variation is also quite remarkable, as indicated both by the large variance in the pronoun condition (see Figure 2) and the frequency data depicted in Table 2.

In the pronoun-true condition, we expected children to be adult-like in most of the trials. The description *Rabbit is sprinkling him* should not be problematic when applied to an other-oriented, two-participant situation. Interestingly, this condition did represent an issue for some of the children in both age groups. The breakdown of the relevant data is in Table 3.

Again, the chi-square test based on the frequency data shows no significant difference between the two age groups. In this condition, the children were expected to perform in an adult-like manner, but in fact some of them (44% in the intermediate group, 34% in the preschooler group)

<sup>14</sup> This preference can also be viewed as an instance of the application of the **Principle of Charity** (see Gualmini et al. 2008). Upon being confronted with a potentially ambiguous sentence, children will prefer the interpretation that makes this sentence true in the given context.

<sup>15</sup> Fritzley and Lee (2003) argue that children are more likely to say *no* than *yes* when confused. If that is the case, then the explanation we provide above is wrong. But it is also possible that older children have learnt how to overcome states of confusion by adopting conscious strategies, like the strategy of staying on the safe side.

**Table 3:** The number of children giving *yes* answers in the given range in Experiment 1, pronoun-true condition

Number of <i>yes</i> answers in the pronoun-true condition	6–5 adult-like	4–2	1–0
Intermediate	8 (50%)	1 (6%)	7 (44%)
Preschooler	8 (44%)	4 (22%)	6 (34%)

rejected these sentences quite consistently. When we asked them to explain their reasons, they consistently replaced the pronoun with a name or a definite noun phrase in their response. So suppose the picture depicts Rabbit sprinkling Christopher Robin. In this condition, the children would typically say that “Rabbit is not sprinkling him, he is sprinkling Christopher Robin”. This means that some of them cannot accommodate an otherwise grammatical pronoun in the absence of a salient (linguistic) discourse antecedent, and they switch to a non-pronominal description. It has been observed independently that children prefer proper names to pronouns in production tasks (see de Villiers et al. 2006), and our results may reflect this preference.<sup>16</sup>

We can conclude from the results of this experiment that Hungarian children show the PIP even at the age of 6–7 years, since we found a remarkable pronoun/reflexive asymmetry in the sentence-picture verification task. Even younger children (aged 4–5) were quite successful when interpreting reflexives, but both age groups performed poorly in the pronoun-false condition. The results do not seem to be far from the classic chance-level performance in this condition, especially in the case of the intermediate group (and some of the preschoolers may have played a yes-strategy for the reasons discussed above). The processing account we discussed in 2.2 provides an adequate explanation for this data, under the assumption that children try to perform reference-set computation in the pronoun-false condition to check whether the coreference reading is available, but

<sup>16</sup> We note here that though the test sentences were presented in isolation in Experiment 1, we did create a larger context for the pictures in the beginning of each individual session (see 3.3). The children watched a picture of each character in the beginning to check if they know them. Then they were told that they were going to watch pictures of these characters in which they are playing and perform different types of activities. We thus created a larger context in which children are prepared for events that are not part of the original stories that they know, but that include characters the children are familiar with and that we listed in the beginning.

they fail because of the processing complexity that this requires. Our results in Experiment 1 thus give further confirmation to Reinhart's (2004; 2006; 2011) account, and provide evidence for the presence of the Pronoun Interpretation Problem in Hungarian.

## 4. Experiment 2: testing sentences with a short lead-in

### 4.1. Hypotheses

As discussed above (2.3), we adopted Spender et al.'s (2009) design in the second experiment, and used a short lead-in to establish a minimal context with a single topic. As Spender et al. (2009) point out, such single topics are very salient in the given discourse and are easily identified as antecedents for pronouns. In fact, they showed that Dutch children's performance in interpreting pronouns improves significantly irrespective of age if the stimulus is presented in this type of discourse context, while the interpretation of reflexives is independent of such discourse factors. Accordingly, we tested the following hypotheses:

**Hypothesis 1** Both age groups will show an improved performance in the pronoun conditions.

**Hypothesis 2** Both age groups will perform at the same level in the reflexive conditions.

### 4.2. Participants

The experimental subjects were again children from the intermediate and the preschooler group of the same local kindergarten, and we tested 16 children from both age groups. We now had to discard only one child from the younger age group for the reason described above (3.2). Table 4 below shows the age and gender details of those children whose performance was evaluated. There was an interval of roughly 4–5 months between the sessions of the two experiments. This explains the increase in average age. Roughly half of the subjects in the first experiment participated in the second experiment, too; and the rest came from among the children who were available at the time of the sessions.

**Table 4:** The participants of Experiment 2

	Boys	Girls	Total	Mean age	Age range
Intermediate	6	9	15	5;0	4;8–5;4
Preschooler	7	9	16	6;11	6;6–7;5

### 4.3. Materials and procedures

Participants again took part in a sentence-picture verification task. The only crucial difference regarding the design of the second experiment was the presence of a lead-in before the target sentence. Consider the two target items below taken from the two experiments:

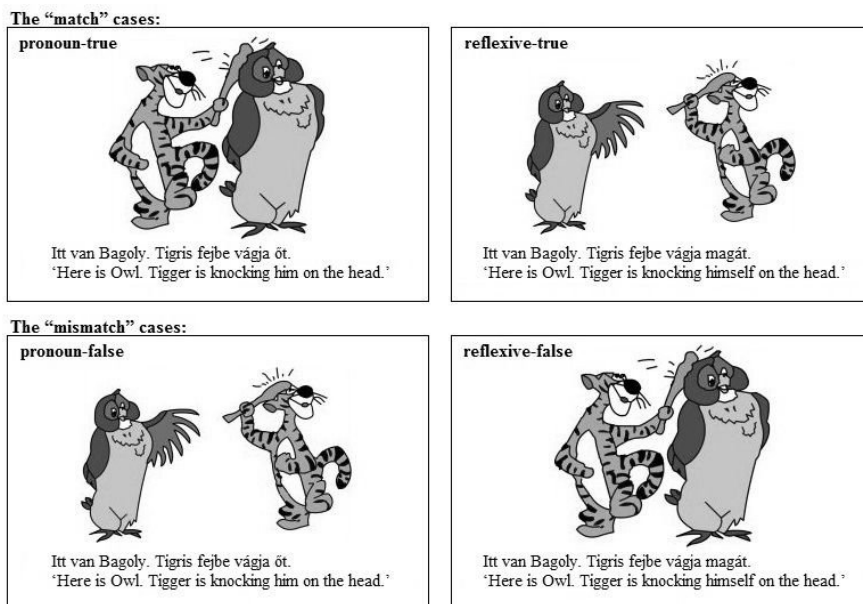
- (13) a. Experiment 1  
 Nyuszi le-locsol-ja ő-t/magá-t.  
 Rabbit.NOM PRT-sprinkle-3SG.DEF-OBJ he-ACC/himself-ACC  
 ‘Rabbit is sprinkling him/himself.’
- b. Experiment 2  
 Itt van Róbert Gida. Nyuszi le-locsol-ja  
 here is Christopher R. Rabbit.NOM PRT-sprinkle-3SG.DEF-OBJ  
 ő-t/magá-t.  
 he-ACC/himself-ACC  
 ‘Here is Christopher Robin. Rabbit is sprinkling him/himself.’

The lead-in introduces a topic that can serve as a natural, non-local antecedent for the pronoun in the pronoun-condition, and which is not expected to interfere with the reflexive condition.

The target sentences contained the same causative verbs as in Experiment 1, however, characters were taken only from three stories, namely, Winnie-the-Pooh, Donald Duck and Little Red Riding Hood. The pictures used in the verification task were drawn in the same manner. We tested exactly the same four conditions, these are illustrated in Figure 3.

The procedures again followed those employed in the first experiment. After an introductory part, the actual sentence-picture verification task consisted of 24 target items and 6 control sentence-picture pairs. The latter contained a non-pronominal definite noun phrase in both the subject and the object position.



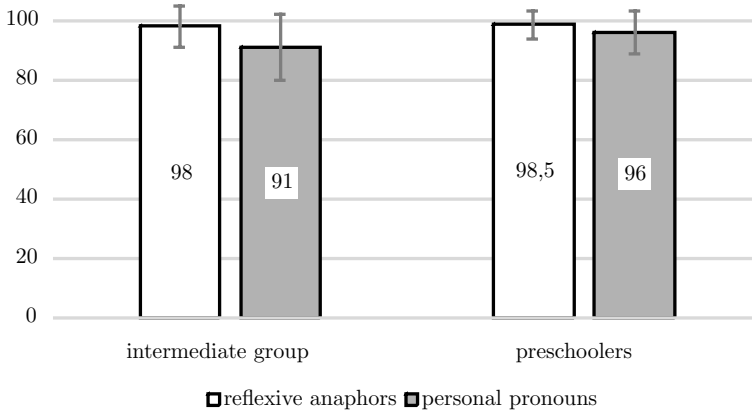


**Figure 3:** The experimental conditions in Experiment 1

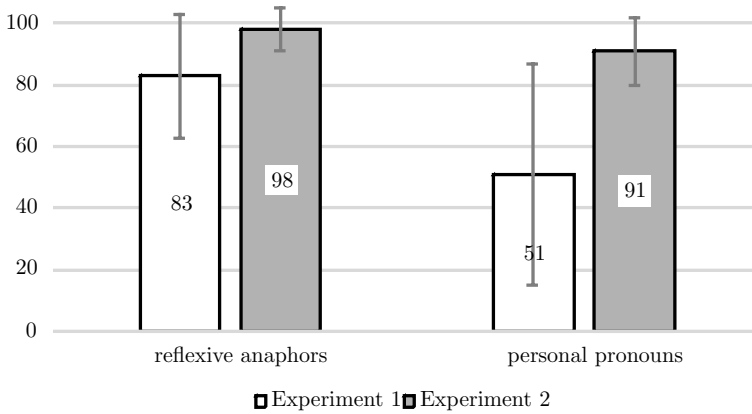
#### 4.4. Results

The results of Experiment 2 are summarized in Figure 4. As expected, using a short lead-in with a single topic resulted in a radical improvement in performance in the pronoun condition to the extent that both age groups came close to adult performance. There is no significant difference between the two age groups in either the pronoun or the reflexive condition. Children in the intermediate group did somewhat better in the reflexive condition than in the pronoun condition, and this difference is significant (paired samples *t*-test,  $t(14) = 2.175$ ,  $p < 0.05$ ). There is no significant difference between the two conditions in the preschooler group.

Figure 5 compares the performance of the intermediate group in the two experiments. The results significantly improved in both conditions across the two experiments (reflexives – independent samples *t*-test (equal variances not assumed):  $t(18) = -2.823$ ,  $p < 0.05$ ; pronouns – independent samples *t*-test (equal variances not assumed):  $t(18) = -4.257$ ,  $p < 0.01$ ), which is surprising, since we did not expect such improvement in the reflexive condition.

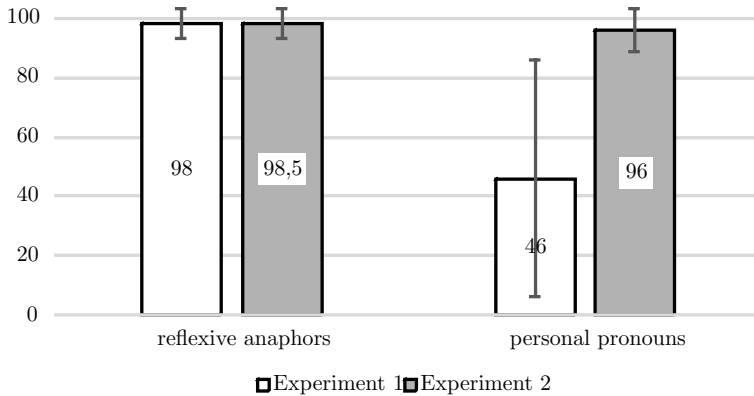


**Figure 4:** Results of Experiment 2: mean correct interpretation of reflexives and pronouns



**Figure 5:** Comparative results of the intermediate group in Experiment 1 and 2: mean correct interpretation of reflexives and pronouns

Finally, Figure 6 compares the results of the preschoolers in the two experiments. There is essentially no change in the reflexive condition, the results of older children were near perfect irrespective of the design. However, the design did have a significant effect on the interpretation of pronouns, and the improvement is obvious (independent samples test (equal variances not assumed):  $t(18) = -5.15$   $p < 0.01$ ).



**Figure 6:** Comparative results of the preschooler group in Experiment 1 and 2: mean correct interpretation of reflexives and pronouns

#### 4.5. Discussion

The primary aim of the second experiment was to test whether a crucial, but minor change in the design, as described by Spender et al. (2009), will effect children's comprehension of pronouns. As mentioned before, Spender et al. (2009) found that the PIP largely disappeared in an experiment carried out with Dutch children when an introductory sentence contained the intended antecedent for the pronoun as a topic. Our results follow the same pattern, namely, children's comprehension of pronouns improved significantly in both age groups with the introduction of a salient discourse antecedent. This is remarkable since it shows that children's comprehension of pronouns relies heavily on contextual clues, and even a very simple way of introducing the topic has a crucial effect on pronoun interpretation. In the second experiment both age groups interpreted pronouns correctly. In fact, their results are very close to the adult pattern (mean correct interpretation is 91% and 96% in the intermediate and the pre-schooler group, respectively). Individual variation (as it is indicated by variance and by the frequency tables below) also became much smaller.

The frequency tables show that both groups achieved an adult-like performance in the pronoun-true condition, since individual children committed at most one mistake out of six cases. The pronoun-false condition was still a bit more difficult for them, but even in the intermediate group 80% showed adult-like performance, i.e., they rejected the description *This is Christopher Robin. Rabbit is sprinkling him* as applied to a reflexive event of Rabbit sprinkling himself. This behaviour clearly shows that introduc-

**Table 5:** The number of children giving *no* answers in the given range in Experiment 2, pronoun-false condition

Number of <i>no</i> answers in the pronoun-false condition	6–5 adult-like	4–2	1–0
Intermediate	12 (80%)	2 (13%)	1 (7%)
Preschooler	15 (94%)	1 (6%)	0

**Table 6:** The number of children giving *yes* answers in the given range in Experiment 2, pronoun-true condition

Number of <i>yes</i> answers in the pronoun-true condition	6–5 adult-like	4–2	1–0
Intermediate	15 (100%)	0	0
Preschooler	16 (100%)	0	0

ing a single topic solves not only the problem that the classic PIP-contexts present, but it also eliminates children's reluctance to accept pronouns in the absence of an accommodating discourse in the pronoun-true condition.<sup>17</sup>

As for the interpretation of reflexives, we did not expect a significant difference in the reflexive conditions, but younger children's performance improved considerably. We do not believe that this improvement is due to design factors, since the Single Topic Condition is not expected to influence judgements in either direction in the reflexive condition.<sup>18</sup> But it may be due to age differences after all, since there was a 4–5 months gap between the two experiments. We suspect that some intermediate children performed worse in Experiment 1 simply because they were younger, but they have matured by the time of Experiment 2.

The results of the two experiments in the pronoun conditions contrast in a much more substantial manner. Children become almost adult-like in both pronoun conditions in Experiment 2. We noted above that some chil-

<sup>17</sup> Remember that the adult control group did not have such issues in Experiment 1 (Figure 2).

<sup>18</sup> Spender et al. (2009) did not find a significant difference either: the Dutch children involved in their experiment provided a correct answer in 86% of the reflexive trials in the Classic Condition, and they did so in the case of 83% of the reflexive trials in the Single Topic Condition. For a description of these conditions, see the discussion in 2.2 above.

dren looked overtly perplexed in Experiment 1 in the pronoun-false condition. We did not experience such behaviour in Experiment 2, and children were generally very quick throughout the second experiment when giving their truth-value judgements.<sup>19</sup> Though these remarks may be anecdotal, they comply with the general results. Children found themselves more at ease in the single topic design, and their responses improved significantly.

We repeat (13b) as (14) to illustrate the pronoun conditions in Experiment 2.

- (14) Itt van Róbert Gida. Nyuszi le-locsol-ja ő-t/magá-t.  
 here is Christopher R. Rabbit.NOM PRT-sprinkle-3SG.DEF-OBJ he-ACC/himself-ACC  
 'Here is Christopher Robin. Rabbit is sprinkling him/himself.'

In the pronoun-true condition, this design creates the minimally coherent discourse that is required for the accommodation of the pronoun. The children uniformly accepted this sentence as the description of a two-participant event in which Rabbit is sprinkling Christopher Robin. In Experiment 1, in the absence of a lead-in, some of the children insisted on the use of the name *Christopher Robin* instead of the pronoun in this case (3.6). We received no such reactions in Experiment 2.

In the pronoun-false condition, (14) is applied to a picture that depicts Rabbit sprinkling himself. Children performed well in this condition, though they were slightly worse than in the pronoun-true condition in Experiment 2 (compare Table 5 and Table 6). This means that only a small residue of the Pronoun Interpretation Problem remains at the comparison of the results in the pronoun-false condition across the two experiments. This is not immediately explained under the processing account of Reinhart (2004; 2006; 2011), since the target transitive construction is the same in the two cases (and only its discourse context changes). Therefore Rule I should apply in exactly the same way in both Experiment 1 and Experiment 2, and we would not expect the improvement that happened in the second experiment. But it did happen. We suggest that what happens in these cases is that children simply do not run Rule I. If a minimally coherent discourse is created with the introduction of an extra-sentential

<sup>19</sup> First we assumed that children themselves could push a button representing fruits for a *yes* answer and a button depicting fruits crossed out for a *no* answer, instead of manipulating toys when rewarding or not rewarding the frog puppet. However, during a pilot experiment we found that children were not able to divide their attention properly between the keyboard and the screen, so in both experiments the experimenter pushed the relevant buttons on the laptop. Consequently, we do not have data about reaction time.

discourse topic, they may simply abide by their knowledge of Principle B, and they are happy to stay with the conclusion that (14) is ungrammatical as a binding configuration. That is, to be more precise, they do not perform the reference-set computation that the last step in the procedure dictated by Rule I requires (5 repeated as 15):

(15) **Covaluation Rule I**

$\alpha$  and  $\beta$  cannot be covalued if

- a.  $\alpha$  is in a configuration to bind  $\beta$ , (namely,  $\alpha$  c-commands  $\beta$ ) and
- b.  $\alpha$  cannot bind  $\beta$  and
- c. the covaluation interpretation is indistinguishable from what would be obtained if  $\alpha$  binds  $\beta$ .

There could still remain some children who try nevertheless. We had three children in the intermediate group and one in the preschooler group who did not perform in an adult-like manner in the pronoun-false condition (Table 5), and one possible explanation is that they ran Rule I.

Such an account implies that Rule I is not an absolute rule for children. They give it a try if no other clues are available to license the pronoun, but they fail because of their processing limitations. If, however, a coherent discourse is created, they may simply opt out of running Rule I. Reinhart (2011) points out in an argument against pragmatic accounts that if children have acquired a pragmatic strategy successfully, then they should perform at the range of 100%. Linking up object pronouns with salient non-local discourse topics is one such strategy, and children performed close to 100% on this task in our Experiment 2. However, Rule I is an innate, absolute rule and its application is therefore obligatory (just like there are no suspensions of the binding principles). What we are suggesting is that it may have a special status in children's grammar, and they may only resort to it in the absence of salient candidates that *could* act as Principle B-compatible antecedents for the object pronoun. This is an issue that requires more consideration, but we leave it at this in this paper, and conclude that our two experiments provide clear evidence that discourse factors play an important role in whether the Pronoun Interpretation Problem surfaces or not.

## 5. Conclusions

In this paper, we reported the results of two child language experiments on the acquisition of Hungarian pronouns and reflexive anaphors. It is well-known that in some languages children have problems when they are asked to interpret pronouns in the context of potential local antecedents, whereas they have no such problems with reflexive anaphors. Following a recent change in terminology, we referred to this effect as the Pronoun Interpretation Problem. Our principal aim was to gain firm evidence whether Hungarian children show this Problem or not.

The two experiments we conducted both involved a picture verification task containing pronouns and reflexives in true and false conditions. The first experiment presented the test sentences in isolation, whereas the second included a short lead-in that provided a natural discourse antecedent for the pronoun. We adopted the design of the second experiment from the Dutch study of Spender et al. (2009), who found that coherent discourse eliminates the Problem, or at least reduces its intensity. Our results are compatible with theirs, and give further support to their claim that children are sufficient users of pragmatic clues.

In particular, we found that Hungarian children had no significant problems with interpreting reflexive anaphors in either design. They had considerable issues with pronouns in the first experiment. In the pronoun-false condition, the classic locus of the Pronoun Interpretation Problem, the group level performance was largely around the chance level. This is especially true of the intermediate group, whereas some preschooler children may have followed a non-adult-like strategy on purpose. Thus Experiment 1 provided evidence that the Pronoun Interpretation Problem is present in Hungarian. However, Experiment 2 provided further evidence that shows that it is very easy to overcome this problem if a minimally coherent discourse is created. Children in both age groups improved to a level that is close to adult performance.

We argued that these results suggest that the Pronoun Interpretation Problem is not due to a lack of grammatical knowledge, in compliance with Reinhart's (2004; 2006; 2011) account. In her model, both the binding principles and the special interface rule that dictates coreference options (Rule I) are innate. Pronouns cannot be locally bound, but they can be locally coreferential by Rule I. Children fail in judgements on local coreference because they cannot perform the computation necessary to decide whether coreference is licensed or not. This results in the guessing pattern that we also saw attested in our Experiment 1. We also tentatively

suggested that in the presence of coherent discourse, children may opt out from running Rule I, and they may perform much better. This happened in our Experiment 2. Further research is needed to strengthen this conclusion and to involve other relevant factors in the investigation, such as the potential effect of the focussing or the *pro*-drop of object pronouns, the inclusion of quantified antecedents, as well as the possible effects of variation in person and number.

### Acknowledgements

We are grateful to the teachers of the kindergarten *Nagyerdei Óvoda* in Debrecen for the assistance we received during the experiments and for their hospitality; and we also thank the parents and the children for their support. We thank Ágnes Lukács and Bence Kas for allowing us to use the pictures they prepared for their experiments on how Hungarian SLI children interpret anaphoric pronominals. We are also thankful to Daniele Panizza (Georg-August-Universität Göttingen) for granting us access to his Pypres toolkit, which we used to conduct our experiments. We are especially grateful to the reviewers of this paper for their extensive comments, and to Katalin É. Kiss for further discussion. Any remaining errors are solely our own.

This research was supported by the Hungarian Scientific Research Fund (OTKA), grant № 111918 (*New approaches in the description of the grammar of Hungarian pronominals*).

### References

- Bánréti, Zoltán. 2006. Neurolingvisztika [Neurolinguistics]. In F. Kiefer and P. Siptár (eds.) *Magyar nyelv* [Hungarian]. Budapest: Akadémiai Kiadó. 653–725.
- Benceze, M. Ildikó. 2014. Kísérleti pragmatika [Experimental pragmatics]. In Cs. Pléh and Á. Lukács (eds.) *Pszicholingvisztika. Magyar pszicholingvisztikai kézikönyv* [Psycholinguistics. Handbook of Hungarian psycholinguistics]. Budapest: Akadémiai Kiadó. 813–858.
- Bloom, Paul, Andrew Barss, Janet Nicol and Laura Conway. 1994. Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language* 70. 53–71.
- Cardinaletti, Anna and Michal Starke. 1995. The tripartition of pronouns and its acquisition: Principle B puzzles are ambiguity problems. In J. N. Beckman (ed.) *NELS 25. Volume 2: Papers from the workshops on language acquisition and language change*. Amherst, MA: University of Massachusetts, GLSA. 1–12.
- Chien, Yu-Chin and Kenneth N. Wexler. 1990. Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* 1. 225–295.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Conroy, Anastasia, Eri Takahashi, Jeffrey Lidz and Colin Phillips. 2009. Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry* 40. 446–486.



- Crain, Stephen and Cecile McKee. 1985. Acquisition of structural restrictions on anaphora. In S. Berman, J.-W. Choe and J. McDonough (eds.) *Proceedings of North East Linguistic Society (NELS) 16*. Amherst, MA: University of Massachusetts, GLSA. 94–110.
- Crain, Stephen and Rosalind Thornton. 1998. *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.
- Czingráber, Márta. 1999. A mondat elemei között fennálló anaforikus viszonyok feldolgozása óvodáskorban, kisiskoláskorban és afáziában [The processing of anaphoric relations between elements of the sentence by kindergarten children, early school children and aphasiacs]. Master's thesis. University of Szeged.
- É. Kiss, Katalin. 2002. *The syntax of Hungarian*. Cambridge: Cambridge University Press.
- Elbourne, Paul. 2005. On the acquisition of Principle B. *Linguistic Inquiry* 36. 333–365.
- Fritzley, V. Heather and Kang Lee. 2003. Do young children always say yes to *yes-no* questions? A metadevelopmental study of the affirmation bias. *Child Development* 74. 1297–1313.
- Grimshaw, Jane and Sara Rosen. 1990. Knowledge and obedience: The developmental status of the binding theory. *Linguistic Inquiry* 21. 187–222.
- Grodzinsky, Yosef and Tanya Reinhart. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24. 69–101.
- Gualmini, Andrea, Sarah Hulsey, Valentine Hacquard and Danny Fox. 2008. The Question–Answer Requirement for scope assignment. *Natural Language Semantics* 16. 205–237.
- Hamann, Cornelia. 2011. Binding and coreference: Views from child language. In J. de Villiers and T. Roeper (eds.) *Handbook of generative approaches to language acquisition*. Dordrecht: Springer. 247–290.
- Hartman, Jeremy, Sudo Yasutada and Ken Wexler. 2012. Principle B and phonologically reduced pronouns in Child English. Paper presented at the Generative Approaches to Language Acquisition, North America (GALANA) 5. University of Kansas, October 11–13, 2012. <http://web.mit.edu/ysudo/www/pdf/GALANA-final.pptx> (2015/07/20)
- Heim, Irene. 1998. Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. In U. Sauerland and O. Percus (eds.) *The interpretive tract (MIT Working Papers in Linguistics 25)*. Cambridge, MA: MIT Press. 205–246.
- Hendriks, Petra and Jennifer Spender. 2005/2006. When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition* 13. 319–348.
- Jakubowicz, Celia. 1984. On markedness and binding principles. In C. Jones and P. Sells (eds.) *Proceedings of North East Linguistic Society (NELS) 14*. Amherst, MA: University of Massachusetts, GLSA. 154–182.
- Ladányi, Enikő, Ágnes Lukács and Bence Kas. in preparation. The role of cognitive control in anaphor interpretation in SLI. Manuscript.
- Lely, Heather K. J. van der and Linda Stollwerck. 1997. Binding theory and grammatical specific language impairment in children. *Cognition* 62. 245–290.
- Pléh, Csaba. 1998. A mondatmegértés a magyar nyelvben [Sentence interpretation in Hungarian]. Budapest: Osiris Kiadó.

- Rákosi, György. 2011. Összetett visszaható névmások a magyarban [Complex reflexives in Hungarian]. *Általános Nyelvészeti Tanulmányok* 23. 351–376.
- Reinhart, Tanya. 1983. Anaphora and semantic interpretation. London: Croom Helm.
- Reinhart, Tanya. 2004. The processing cost of reference-set computation: Acquisition of stress shift and focus. *Language Acquisition* 12. 109–155.
- Reinhart, Tanya. 2006. *Interface strategies*. Cambridge, MA: MIT Press.
- Reinhart, Tanya. 2011. Processing or pragmatics? Explaining the coreference delay. In E. Gibson and N. Perlmutter (eds.) *The processing and acquisition of reference*. Cambridge, MA: MIT Press. 157–196.
- Reinhart, Tanya and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24. 657–720.
- Rooryck, Johann and Guido Vanden Wyngaerd. 2011. *Dissolving binding theory*. Oxford: Oxford University Press.
- Rooryck, Johann and Guido Vanden Wyngaerd. 2015. Morphological transparency and the Delay of Principle B Effect. *Lingua* 155. 121–139.
- Safir, Ken. 2004. *The syntax of anaphora*. Oxford: Oxford University Press.
- Spenader, Jennifer, Erik-Jan Smits and Petra Hendriks. 2009. Coherent discourse solves the Pronoun Interpretation Problem. *Journal of Child Language* 36. 23–52.
- Thornton, Rosalind and Kenneth Wexler. 1999. Principle B, VP-ellipsis, and interpretation in child grammar. Cambridge, MA: MIT Press.
- de Villiers, Jill G., Jacqueline Cahillane and Emily Altreuter. 2006. What can production reveal about Principle B? In K. U. Deen, J. Nomura, B. Schulz and B. D. Schwartz (eds.) *The proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition, North America (University of Connecticut Occasional Papers in Linguistics 4)*. Storrs: University of Connecticut, Department of Linguistics. 89–100.
- Wexler, Kenneth and Yu-Chin Chien. 1985. The development of lexical anaphors and pronouns. *Papers and Reports on Child Language Development* 24. 138–149.