

# The disunity of Principle B Effects

Giuseppe Varaschin

## Abstract

It is a typologically well-attested generalization that, in languages that have simple personal pronouns (p-pronouns) and dedicated reflexive forms, the former are avoided when the purpose is to signal semantic identity between the coarguments of a predicate (Faltz, 1985; Comrie, 1999; Levinson, 2000; Mattausch, 2007; Haspelmath, 2008; Reuland, 2011; Volkova & Reuland, 2014). English is one such language:

- (1) a. \*Susan<sub>1</sub> praised her<sub>1</sub>.
- b. \*Marta<sub>1</sub> voted for her<sub>1</sub>.
- c. \*Bobby<sub>1</sub> thinks of him<sub>1</sub> first, then others.
- d. \*Joanne<sub>1</sub> forgot to include her<sub>1</sub> in the guest list.

Most researchers within HPSG follow Mainstream Generative Grammar in the assumption that Principle B of the Binding Theory suffices to explain the kinds of disjointness effects exemplified above (Chomsky, 1981; Pollard & Sag, 1994; Manning & Sag, 1998; Runner & Kaiser, 2005). The following is a standard statement of Principle B within HPSG:

- (2) PRINCIPLE B: A p-pronoun must be locally o-free.

However, the idea that a single syntactic constraint provides a universally valid account for disjointness effects involving p-pronouns does not hold up to a wider survey of the data. As an illustration, consider how the judgments for the English examples in (1) only partially overlap with those for the corresponding Brazilian Portuguese (BP) examples in (3) (Lemle, 1985; Galves, 1986; Grolla, 2011; Grolla & Bertolino, 2011):

- (3) a. \*A Susan<sub>1</sub> louvou ela<sub>1</sub>.  
      the Susan praised her
- b. \*A Marta<sub>1</sub> votou nela<sub>1</sub>.  
      the Marta voted in-her
- c. O Roberto<sub>1</sub> pensa primeiro nele<sub>1</sub>, depois nos outros.  
      the Roberto thinks first on-him, then on-the others
- d. A Joana<sub>1</sub> esqueceu de incluir ela<sub>1</sub> na lista de convidados.  
      the Joana forgot to include her in-the list of guests

The data in (3) show that p-pronouns in BP need not always be o-free. This presents a major puzzle for approaches based exclusively on (2): theorists can neither affirm that p-pronouns in BP are run-of-the mill Principle B-abiding forms like their English counterparts, nor that these items are completely exempt from Principle B effects. A similar dual behavior is found in connection to p-pronouns in Middle English (Faltz, 1985), as in (4), Frisian (Reinhart & Reuland, 1993), as in (5), and French (Zribi-Hertz, 1995), as in (6):<sup>1</sup>

- (4) a. He<sub>1</sub> cladde hym<sub>1</sub> as a poure laborer.  
      he dressed him as a poor laborer  
      ‘He<sub>1</sub> dressed him(self)<sub>1</sub> as a poor laborer.’
- b. \*Hym<sub>1</sub> he<sub>1</sub> hynge.  
      him he hanged  
      ‘He<sub>1</sub> hanged him<sub>1</sub>.’
- (5) a. Max<sub>1</sub> wasket him<sub>1</sub>.  
      Max washes him  
      ‘Max<sub>1</sub> washes him(self)<sub>1</sub>.’

---

<sup>1</sup>The judgment in (4-b) is a hypothesis motivated by the unexpected absence of such data in corpora: locally o-bound *hym* (‘him’) only seems to occur with a restricted class of predicates, from which *hynge* (‘hang’) is excluded (Faltz, 1985). More on this below. The possibility of local o-binding of p-pronouns in French seems to be restricted to cases where the p-pronouns occur within PPs (Reuland, 2011). However, this is still problematic for Principle B because the PPs in (6) are not the kinds of PPs that introduce their own ARG-STR lists.

- b. \*Max<sub>1</sub> hatet him<sub>1</sub>.  
 Max hates him  
 ‘Max<sub>1</sub> hates him<sub>1</sub>.’
- (6) a. Pierre<sub>1</sub> est fier de lui<sub>1</sub>.  
 Pierre is proud of him  
 ‘Pierre<sub>1</sub> is proud of him(self)<sub>1</sub>.’
- b. \*Pierre<sub>1</sub> est jaloux de lui<sub>1</sub>.  
 Pierre is jealous of him  
 ‘Pierre<sub>1</sub> is jealous of him<sub>1</sub>.’

There is no syntactic generalization that distinguishes the (a) and (b) cases in a general way. Rather, the difference seems to be related to a pragmatic property of the predicates where the p-pronouns appear: the more stereotypical a non-reflexive interpretation is for a predicate, the less acceptable are the locally o-bound p-pronouns within it (Zribi-Hertz, 1995).<sup>2</sup>

Even for Modern English, the idea that a principle like (2) exhausts the range of disjointness effects involving p-pronouns is problematic. Since it is generally assumed that the indices borne by the coargument NPs in (7) are not identical, but merely assigned to the same entity, Principle B cannot be relied upon to rule out semantic identity in a parallel case like (8), assuming an assignment  $g$  where  $g(1) = g(2)$  (Reinhart, 1983; Pollard & Sag, 1994):

- (7) a. If everybody voted for John<sub>1</sub>, surely JOHN<sub>2</sub> must have voted for him<sub>1</sub>.  $g(1) = g(2)$   
 b. At the end of the day, only Donald<sub>1</sub> voted for him<sub>2</sub>.  $g(1) = g(2)$
- (8) \*Marta<sub>1</sub> voted for her<sub>2</sub>.  $g(1) = g(2)$

The examples in (3)-(8) strongly suggest that disjointness effects typically attributed to Principle B do not correspond to a unified phenomenon. I propose that the responsibility for accounting for these effects should be distributed into three separate factors:

- (i) A language-specific lexical property of predicates that prohibits sharing the INDEX values of their p-pronoun complements with any o-commanding coarguments.
- (ii) A pragmatically-grounded constraint on the morphosyntactic encoding of reflexive predications (Faltz, 1985; Comrie, 1999; Levinson, 2000; Mattausch, 2007).
- (iii) A preference for expressing semantic identity by means of coindexation instead of by assigning distinct indices to the same referent, unless the latter yields an interpretive effect distinct from coindexation (Reinhart, 1983; Krifka, 2018).

The principle in (i) basically corresponds to the HPSG version of Principle B. However, instead of viewing Principle B as a universal syntactic constraint as HPSG typically does, (i) proposes to push the generalization it expresses into the lexicon. This potentially makes it easier to account for languages that don’t exhibit the effects of Principle B in the same way Modern English does. Since the lexicon contains a large bulk of what is learnable and variable in language, it will come as no surprise that many languages might come to lack a general *syntactic* constraint against the local o-binding of p-pronouns (cf. (3)-(6)).

One way to implement (i) is by using HPSG’s hierarchy of lexical types to encode different kinds of local binding requirements in terms of the anaphoric realizations of a predicate’s arguments. For English, we could divide the lexical class associated with transitive predicates into two disjoint subsorts, defined by whether they take non-p-pronouns (*nppro*) or p-pronouns (*ppro*) as complements:

- (9) a. *trans-pred-nppro-lxm*  $\Rightarrow$  [ARG-STR  $\langle$ NP $\rangle \oplus$  *nelist*(NP:*nppro*)]  
 b. *trans-pred-ppro-lxm*  $\Rightarrow$  [ARG-STR  $\langle$ NP<sub>1</sub>, (NP<sub>2</sub>) $\rangle \oplus$  (NP:*ppro*<sub>-1 $\wedge$ -2</sub>)]

Only subsorts of *trans-pred-ppro-lxm* are able to accept p-pronouns as complements. But, as a consequence of (9-b), all such predicates will disallow coindexing between the p-pronoun and the NPs that o-command it in the ARG-STR list. In order to capture the fact that any predicate

<sup>2</sup>The notion of stereotypical (non-)reflexivity is familiar from the functionalist literature on anaphora (Faltz, 1985; Comrie, 1999; Ariel, 2008; Levinson, 2000; Haspelmath, 2008). It is based on the intuition that reflexivity is less usual for some predicates (e.g. *hang*, *jealous*) than for others (e.g. *dress*, *proud*). This is plausibly reflected in frequency of reflexive use: given a large enough corpus of utterances, stereotypically non-reflexive predicates like *kiss* will occur less often with reflexive interpretations than predicates like *proud* or *shave*, which are more stereotypical (or neutral) with respect to reflexivity. See Ariel (2008) and Haspelmath (2008) for some frequency counts that confirm this prediction.

that can take a non-p-pronoun complement has a counterpart that can take a p-pronoun complement (with the aforementioned consequences for the INDEX values of its arguments) we must formulate a lexical rule that maps lexemes satisfying (9-a) into lexemes satisfying (9-b), as Jacobson (2007) does in her categorial grammar account.

What we have to say for languages that do allow locally o-bound p-pronouns (e.g. BP, Middle English and Frisian) is that the constraint they impose on transitive predicates contains no restriction on ARG-STR values (something similar works for French PP complements):

- (10) *trans-pred-lxm*  $\Rightarrow$  [ARG-STR  $\langle$ NP $\rangle \oplus \textit{nelist}(\textit{NP})]$

Since (10) does not specify the CONTENT values for the NPs in the ARG-STR, no specific lexical class constraint for licensing p-pronoun complements is needed. This means that there is nothing to block coindexation between p-pronouns and their o-commanding coarguments. This explains why locally o-bound p-pronouns are acceptable in (3-c)-(3-d), (4-a) and (5-a).

The cases where locally o-bound p-pronouns are *not* acceptable in BP, Middle English, Frisian and French (cf. (3-a)-(3-b), (4-b), (5-b) and (6-b)) are not handled by a syntactic constraint like (i), but by the constraint suggested in (ii), whose full statement is given below:

- (11) CONSTRAINT ON REFLEXIVE PREDICATIONS (CRP):  
 If the CONTENT|RELS value of a *synsem* object *S* contains a reflexive elementary predication *R* and *R* is not stereotypically reflexive, then *S* must be reflexive-marked.  
 a. *R* is reflexive iff the values for two ARG attributes of *R* are structure-shared;  
 b. *S* is reflexive-marked iff a member of *S*'s ARG-STR list is NP:*refl*.

The CRP is similar to the Condition B of Reinhart & Reuland's (1993) Reflexivity Theory. Unlike Reinhart & Reuland's principle, however, (11) should not be seen a primitive, but as a consequence of a universal pragmatic principle that associates unmarked forms with stereotypical interpretations – namely, Levinson's (2000: 37) I(NFORMATIVENESS)-PRINCIPLE:

- (12) I-PRINCIPLE: What is expressed simply is stereotypically exemplified.

The idea is that, since p-pronouns are simple unmarked forms (in contrast to reflexives), they trigger an I-PRINCIPLE-based inference to a stereotypical interpretation for each *synsem* object within which they occur. This means that p-pronouns can only express a reflexive predication *R* in *synsem* objects for which *R* being reflexive is stereotypical. If a non-reflexive interpretation is stereotypical, p-pronouns will trigger non-reflexivity. The only way to signal reflexivity in such cases is by resorting to specialized reflexive-marking.

Logically, the CRP is a conditional statement of the form *If A and B, then C*, where C is the reflexive-marking requirement. If either one of the conjuncts of the antecedent (A or B) is false, reflexive-marking is not necessary. This gives us basically two logically possible scenarios where a locally o-bound p-pronoun may avoid violating the CRP:

- (13) a. When the elementary predication *R* is stereotypically reflexive.  
 b. When the elementary predication *R* is not reflexive.

The scenario in (13-a) covers examples like (3-c), (4-a), (5-a) and (6-a). The scenario in (13-b) covers cases where the meaning of the p-pronoun is not literally identical to that of its antecedent, but is shifted to denote a proxy of the latter (Safir, 2004; Varaschin, 2020). This is what happens in (3-d) above: the elementary predication introduced by *incluire* ('include') expresses a relation between Joana and a proxy of Joana (namely, *her name*). It is also what happens in the BP example (14), where *ela* ('her') is interpreted as a visual image of Marta:

- (14) A Marta<sub>1</sub> viu ela<sub>1</sub> cantando na TV.  
 the Marta saw her singing on-the TV.  
 'Marta<sub>1</sub> saw her(self)<sub>1</sub> singing on TV.'

Since the CRP is grounded in the I-PRINCIPLE, it should be universal. So we expect to see some of its effects in English constructions that are exempt from the effects of (9-b). This is the case of locative PPs, as (15) shows (Chomsky, 1981; Reinhart & Reuland, 1993):

- (15) a. Bobby<sub>1</sub> rolled the carpet over him<sub>1</sub>.  
 b. April<sub>1</sub> put the book near her<sub>1</sub>.

Despite being exempt from (9-b) (because locative PPs only have a single element in their

ARG-STR lists), when the *synsem* object that corresponds to the PP does contain a reflexive elementary predication among the values of CONTENT|RELS, CRP demands reflexive-marking:

- (16) a. Bobby rolled the carpet<sub>1</sub> over itself<sub>1</sub> / \*it<sub>1</sub>. (Reinhart & Reuland, 1993, 687)  
 b. Bobby<sub>1</sub> stepped on himself<sub>1</sub> / \*him<sub>1</sub>. (Reinhart & Reuland, 1993, 688)

Another context in which English p-pronouns are exempt from (9-b) are conjoined NPs. This happens because a p-pronoun that occurs inside an NP is not a direct member of the ARG-STR list of its predicate (Reinhart & Reuland, 1993). In spite of this, conjoined NPs do seem to be contexts in which p-pronouns are sensitive to the CRP:

- (17) a. Joanne<sub>1</sub> forgot to include Larry and her<sub>1</sub> in the guest list.  
 b. ??Joanne<sub>1</sub> forgot to praise Larry and her<sub>1</sub> at the party.  
 (18) a. Paul<sub>1</sub> thinks of Amy and him<sub>1</sub> first, then others.  
 b. ??Paul<sub>1</sub> poked Amy and him<sub>1</sub> first, then others.

In (17-a), the p-pronoun refers to a proxy of its antecedent. As a consequence, no reflexive elementary predication is expressed. If (18-a), we do find a reflexive predication (under a distributive reading where Paul is both the AGENT and the THEME of *think*), but this is one for which reflexivity is stereotypical. The predications implied in (17-b)/(18-b), on the other hand, are all non-stereotypically reflexive. Since these instances of non-stereotypical reflexivity are not accompanied by reflexive-marking of the *synsem*, they are excluded by the CRP.

Note that neither (9-b) nor (11) exclude cases where the INDEX values of two coargument NPs are type-identical but token-distinct, as in *He<sub>1</sub> voted for him<sub>2</sub>*. As we saw in (7) above, it is possible that structures like these may be used to signal coreference in peculiar contexts. This particular mode of achieving coreference is not mediated by identity of indices, but established on purely pragmatic grounds by anchoring distinct indices to the same referent.

It is good, therefore, that this kind of contextually determined coreference is not ruled out by (9-b) and (11). This means, however, that something other than (9-b) and (11) has to explain why we do not get coreference in cases like (8). This is where the factor in (iii) above comes into play. Following Reinhart (1983), Reuland (2011), Krifka (2018) and others, I argue that (iii) can be cashed out as a consequence of Levinson's (2000) M(ANNER)-PRINCIPLE:

- (19) M-PRINCIPLE: Avoid prolix, obscure or marked expressions without reason.

If we assume that a sign which grammatically encodes an interpretation  $\mathcal{I}$  is more explicit than one that does not encode  $\mathcal{I}$ , but allows the speaker to contextually infer  $\mathcal{I}$ , the structure in (20-a) will count as more explicit than (20-b). Therefore, the M-PRINCIPLE will preempt the latter unless the speaker has some overriding reason to avoid (20-a).

- (20) a. Marta<sub>1</sub> voted for herself<sub>1</sub>.  
 b. \*Marta<sub>1</sub> voted for her<sub>2</sub>.  $g(1) = g(2)$

More generally, speakers will only opt for assigning the same referent to distinct indices if there is some particular motivation for not using a plain coindexed structure – e.g. if there is some specific interpretive effect attainable solely by a non-coindexed variant. This effect can be a shift in attitude towards the referent or a desire to avoid a bound-variable interpretation, which is the default whenever a pronoun is o-commanded by its antecedent.

The latter is precisely what happens in cases like (7), according to Reinhart (1983). In (7-b) (*Only Donald<sub>1</sub> voted for him<sub>2</sub>*), for example, pragmatic coreference with distinct indices is possible if the speaker specifically intends to convey that the property that only Donald possesses is the property of having voted for Donald ( $\lambda x. x$  voted for Donald) and not the property of having voted for oneself ( $\lambda x. x$  voted for  $x$ ). This last property would be the one obtained in a coindexed structure, since it reflects a bound-variable reading of the p-pronoun.

To summarize, then, we find that we need three factors to take up the explanatory burden previously attributed to Principle B: (i) a lexical constraint on the argument structure of predicates; (ii) a constraint on the morphosyntactic encoding of semantically reflexive elementary predications derivable from the I-PRINCIPLE, (iii) a preference for coindexing derivable from the M-PRINCIPLE. Unlike the traditional Principle B in (2), none of these factors is a *syntactic* universal. It is the occasional absence of (i) and the universal presence of (ii) and (iii) that explain the peculiar anaphora patterns we observe across different languages.

## References

- Ariel, Mira. 2008. *Pragmatics and grammar*. Cambridge: Cambridge University Press.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Comrie, Bernard. 1999. Reference-tracking: description and explanation. *STUF-Language Typology and Universals* 52(3-4). 335–346.
- Faltz, Leonard M. 1985. *Reflexivization: A study in universal syntax*. New York, NY: Garland.
- Galves, Charlotte. 1986. A interpretação “reflexiva” do pronome no português do Brasil. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada* 2(2).
- Grolla, Elaine. 2011. The acquisition of contrastive and non-contrastive anaphoric forms in Brazilian Portuguese. In *Selected proceedings of the 4th Conference on Generative Approaches to Language Acquisition North America*, 78–89.
- Grolla, Elaine & Karina G Bertolino. 2011. A proforma ‘ele’ com antecedente local em português brasileiro adulto e infantil. In *Estudos da linguagem: casamento entre temas e perspectivas*, Joao Pessoa: Editora Universitária da UFPB.
- Haspelmath, Martin. 2008. A frequentist explanation of some universals of reflexive marking. *Linguistic Discovery* 6(1). 40–63.
- Jacobson, Pauline. 2007. Direct compositionality and variable-free semantics: the case of “Principle B” effects. In Chris Barker & Pauline Jacobson (eds.), *Direct compositionality*, Oxford: OUP.
- Krifka, Manfred. 2018. A direct compositionality approach to Condition C effects under reconstruction and their exceptions. In Manfred Krifka & Mathias Schenner (eds.), *Reconstruction effects in relative clauses*, vol. 75, 35. Walter de Gruyter GmbH & Co KG.
- Lemle, Miriam. 1985. Pronomes, anáforas, zero: observações sobre uma mudança lingüística. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada* 1(1/2).
- Levinson, Stephen C. 2000. *Presumptive meanings*. Cambridge, MA: MIT Press.
- Manning, Christopher D & Ivan A Sag. 1998. Argument structure, valence, and binding. *Nordic Journal of Linguistics* 21. 107–144.
- Mattausch, Jason. 2007. Optimality, bidirectionality & the evolution of binding phenomena. *Research on language and computation* 5(1). 103–131.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press and CSLI Publications.
- Reinhart, Tanya. 1983. *Anaphora and semantic interpretation*. Chicago, IL: University of Chicago Press.
- Reinhart, Tanya & Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24(4). 657–720.
- Reuland, Eric. 2011. *Anaphora and language design*. Cambridge, MA: MIT Press.
- Runner, Jeffrey T & Elsi Kaiser. 2005. Binding in picture noun phrases: Implications for binding theory. In Stefan Müller (ed.), *Proceedings of the 12th international conference on Head-Driven Phrase Structure Grammar*, 594–613. Stanford, CA: CSLI Publications.
- Safir, Ken. 2004. *The syntax of anaphora*. New York, NY: Oxford University Press.
- Varaschin, Giuseppe. 2020. Anti-reflexivity and logophoricity: an account of unexpected reflexivization contrasts. *Glossa: a journal of general linguistics* 5(1).
- Volkova, Anna & Eric Reuland. 2014. Reflexivity without reflexives? *The Linguistic Review* 31(3-4). 587–633.
- Zribi-Hertz, Anne. 1995. Emphatic or reflexive? on the endophoric character of french lui-même and similar complex pronouns. *Journal of linguistics* 333–374.