

Remarks on Binding Theory

Carl Pollard

Ohio State University

Proceedings of the 12th International Conference on
Head-Driven Phrase Structure Grammar

Department of Informatics, University of Lisbon

Stefan Müller (Editor)

2005

Stanford, CA: CSLI Publications

pages 561–577

Pollard, Carl. 2005. Remarks on Binding Theory. In Stefan Müller (ed.), *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, 561–577. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2005.32.



Abstract

We propose some reformulations of binding principle A that build on recent work by Pollard and Xue, and by Runner et al. We then turn to the thorny issue of the status of indices, in connection with the seemingly simpler Principle B. We conclude that the notion of index is fundamentally incoherent, and suggest some possible approaches to eliminating them as theoretical primitives. One possibility is to let logical variables take up the explanatory burden borne by indices, but this turns out to be fraught with difficulties. Another approach, which involves returning to the idea that referentially dependent expressions denote identity functions (as proposed, independently, by Pollard and Sag and by Jacobson) seems to hold more promise.

1 Introduction

As formulated by Chomsky (1986), binding theory (hereafter BT) constrained **indexings**, which were taken to be assignments of **indices** to the NPs in a phrase. What an index was was irrelevant; what mattered was that they partitioned all the NPs in a phrase into equivalence classes. Phrases, in turn, were taken to be trees of the familiar kind and the binding constraints themselves were couched in terms of tree-configurational notions such as **government**, **c-command** (or **m-command**), **chain**, and **maximal projection**. In the early 1990's, numerous studies (Everaert, 1991; Hellan, 1991; Pollard and Sag, 1992; Pollard and Sag, 1994; Reinhart and Reuland, 1991; Reinhart and Reuland, 1993) converged on the view that a wide range of facts at odds with Chomsky's BT became explicable if the binding constraints were reformulated in terms of the argument structures of the predicates rather than tree configurations. Additionally, many of these same investigators and others (Sells, 1987; Zribi-Hertz, 1989; Baker, 1995; Pollard and Xue, 1998; Pollard and Xue, 2001; Golde, 1999; Runner et al., 2002) recognized that a distinction had to be drawn between (1) occurrences of referentially dependent elements subject to syntactically characterizable constraints on their (linguistic) antecedents, and (2) occurrences subject to interpretive constraints couched in terms of such discorsal/information-structural notions as logophoricity, discourse prominence, and contrastiveness.¹ Following a common usage, we will limit the term "BT" to constraints of the first kind, and speak of occurrences of referentially dependent expressions which are subject only to the second kind of constraint as **exempt** from BT.

[†]For helpful discussion, I would like to thank the participants in the Lisbon Binding Theory Workshop, especially Yusuke Kubota, Jeff Runner, and Ping Xue. The research reported here was supported by a grant from the National Science Foundation.

¹It is often assumed that there is lexical ambiguity between the first type and the second, but another possibility is that expressions which have been analyzed in this way are not ambiguous, but rather are subject to syntactic constraints in some environments and to discorsal/information-structural ones in other environments. See Pollard and Xue 1998 and 2001 for discussion.

The following remarks are organized as follows. In section 2, we review a (somewhat dated) HPSG-based formulation of Principles A and B (Pollard and Sag 1992, 1994), which we will use as a point of departure. Section 3 examines two recent but mutually inconsistent refinements of Principle A and seeks a synthesis to resolve the inconsistency. In section 4, we turn to Principle B, which is usually considered simpler than Principle A since there are not the troublesome notions of reflexivity (or reciprocity) and exemption to contend with. But, we will argue, the simplicity is illusory, because it is with Principle B that we are forced to confront in its purest form the even more vexing question of just what exactly it is that BT constrains. Section 5 draws some tentative conclusions and suggests some directions for future investigation.

2 Pollard and Sag’s Reformulation of Principles A and B

2.1 HPSG Background

Pollard and Sag’s reformulations of Chomsky’s (1986) first two BT principles are couched not in tree-configurational terms but rather in terms of **valence**, which is a certain technical embodiment of the notion of syntactic argument structure. Words are assumed to select their subjects, specifiers, and complements via **valence features** (SUBJ, SPR², and COMPS respectively). This applies not only to verbs (including auxiliaries) but also to other argument-taking words, such as nouns with possessive determiners or PP complements, and predicative adjectives and prepositions, including the lexical heads of absolutive sentential modifiers and so-called “reduced relatives” (postnominal predicative modifiers).

Except for dummy *it* and *there*, NPs and “case-marked” PPs (ones with semantically vacuous prepositions) in English are assumed to have an **index**. These include not just phonetically realized elements, but also the HPSG analogs of inaudible elements such as PRO (e.g. unexpressed subjects of VP complements) and syntactic variables (gaps in unbounded dependencies).³

Every indexed element belongs to one of the three **reference types: r-pronoun, p-pronoun, or non-pronoun**. R-pronouns include overt reflexives and reciprocal *each other*; p-pronouns include ordinary nonreflexive definite pronouns; and all other overt indexed NPs, including names, relative and interrogative “pronouns”, and other NPs headed by common nouns, are non-pronouns. Some standard analytic assumptions are the following: (1) Any case-marked PP has the same reference type and the same index as the object of the preposition; (2) in an unbounded dependency, the trace has the same reference type and the same index as the filler; and (3) in raising (to subject or object), the unrealized complement subject has the same reference type and the same index as the controller.

²The notion of specifier employed here is different from the GB notion of the same name. Here determiners (including possessive ones) are analyzed as specifiers of nouns and comparative gradables, and degree phrases as specifiers of noncomparative gradables; but subjects are distinct from specifiers, and fillers (“extracted” phrases) are not analyzed as valents at all.

³For present purposes, we will speak as if gaps are analyzed as phonetically null constituents (traces), but nothing hinges on this.

The valents of a word have an abstract linear order (possibly different from the temporal order of their phonetic realizations) called the **obliqueness** order. The centrally important notion of **local o-command** is defined in terms of the obliqueness order as in (1):

(1) Local o-command

For *X* and *Y* two indexed valents (subjects, specifiers, or complements) of a word, *X* **locally o-commands** *Y* just in case it precedes *Y* in the obliqueness order of that word's valents.

At least for English, the following analytic assumptions have been standard. First, indexed subjects and specifiers locally o-command indexed complements; in particular, a possessive determiner of a noun locally o-commands a nonpredicative PP complement of the noun, as in so-called "picture NPs" (PNPs) such as *Mary's picture of herself*. And second, if both subject and specifier are present (e.g. *Mary considers John her ally*), the subject (here, *John*) locally o-commands the specifier (here, *her*).

Pollard and Sag's reformulations of Principles A and B then run as follows:

(2) Binding Theory for English (Pollard and Sag 1992, 1994)

- a. Principle A: Every locally o-commanded r-pronoun is coindexed with one of its local o-commanders.
- b. Principle B: Every p-pronoun is coindexed with none of its local o-commanders.

It is important to note that r-pronouns which are *not* locally o-commanded are *not* required by this formulation of BT to be coindexed with anything else; though their interpretation *is* assumed to be subject to other, nonsyntactic, constraints of a discursal or information-structural nature. Such r-pronouns are said to be **BT-exempt**, or simply **exempt**.

Some exempt positions for r-pronouns are listed in (3):

(3) Some Exempt positions for r-pronouns

- a. subjects of nonfinite ("small") clauses
- b. objects of verbs with dummy *it* subjects
- c. (possessive) determiners of nonpredicative NPs
- d. PP complements of nonpredicative NPs without possessive determiners.

By comparison, Chomsky's (1986) form of Principle A wrongly requires such r-pronouns ("anaphors" in his terminology) to be "A-bound" i.e. coindexed with a c-commanding NP in an argument position within a certain "governing category" (specifically, the least maximal projection containing a subject, the r-pronoun, and the r-pronoun's governor). For a nonexhaustive list of kinds of examples where the Pollard-Sag account compares favorably with Chomsky's, see Pollard and Sag 1994, p. 245.

In spite of the many technical and empirical differences between the Pollard-Sag BT and Chomsky's (see Pollard and Sag 1992 and 1994 for extensive discussion), there are also many striking similarities, including the following. First, in a candidate structure being considered with respect to BT-compatibility, each nondummy NP has associated with it something called its index. Second, the structures contain substructures (either subtrees or sub-feature-structures) corresponding not just to overt NPs but also to controllers (PRO, null pronouns (pro), gaps (*variables* in the GB sense), and raised NPs (NP-trace). Third, overt controllers, "raised" constituents, and "wh-moved" constituents (fillers in HPSG, non-null heads of \bar{A} -chains in GB) are coindexed with the corresponding "abstract" elements (PRO, pro, variable, or NP-trace in GB; a member of the list value of some valence feature in HPSG⁴). Fourth, in cases where a quantified NP (hereafter, QNP) semantically binds a pronoun or reflexive (in the sense that in a standard logical translation, the logical determiner of the QNP logically binds two logical variable instances, one from the QNP itself and one from the pronoun/reflexive), the QNP and the pronoun/reflexive are coindexed; and this holds true even if the QNP does not o-command/c-command the pronoun/reflexive, as in examples such as the following:

- (4) a. The first dollar he ever earned is the most treasured possession of many a successful entrepreneur.
 b. Some crank in every little midwestern would like to burn it to the ground.

Fifth, coreference need not occasion coindexing, as illustrated in the following examples:

- (5) a. He's the man that shot Liberty Valance. [speaker pointing at Black Bart]
 b. While he was suffering from amnesia, Nixon didn't realize that he was actually Richard Nixon.

Sixth, by virtue of Principle A, some r-pronoun/anaphor occurrences (exactly which ones depending on the theory) are required to bear *the same* index as certain other NPs. And seventh, certain pronoun/pronominal occurrences (again, exactly which ones depending on the theory) are required to bear indices which are *distinct* from those of certain other NPs.

It is striking that two theories formulated within frameworks that differ so dramatically in terms of their methodological assumptions and theoretical primitives should agree on so much. So striking, in fact, that one might well suspect they are two theories about the same things, things which both theories call **indices**. But what are these indices that the two BTs seem to bear about? In the following subsection, we review what HPSG says about this, but we will return to this question from a less theory-bound perspective in section 4.

2.2 Indices in HPSG

On the Pollard-Sag account, indices are not just integers (or other unique identifiers) assigned to NP nodes in trees as they are in GB. Rather, for each nondummy

⁴Or, in recent HPSG, a member of the list value of the ARGSTRUC feature.

NP (or case-marked PP), irrespective of its feature-structure type (sign, synsem, or local), there is a feature path terminating with CONTENT|RESTIND|INDEX leading to a substructure of type index, which in turn bears a set of features usually called *agreement* features (usually, PERSON, NUMBER, GENDER). This is the case no matter whether the index-bearing element is a QNP (e.g. *every boy*), a name (e.g. *Kim*), a pronoun/reflexive, or one of the “abstract” ARGSTRUC elements that does not correspond directly to a realized sign.

Let us consider some of the cases. (1) In the case of a QNP, the index occurs in the feature-structural representation of the logical quantifier in the position corresponding to that of the first x in $\forall x(\text{boy}'(x))$ in a restricted-quantification logical representation. (2) For a name, the index is playing a role in the CONTENT value essentially like the one that would be played by a logical constant (say, in a translation into intensional logic). (3) For a bound pronoun, reflexive, pro, PRO, or trace, the index is playing a role analogous to the one that would be played by an occurrence of a logical variable in an argument position of a predicate in a logical translation. (4) And for a deictic or logophoric pronoun, the index is playing a role analogous to that of a logical parameter/indeterminate (i.e. a free variable whose reference is fixed by the utterance context). What is problematic here is that in the kind of semantics that 1990's-style HPSG CONTENT values are supposed to be modelling (i.e. west-coast-style situation semantics), there no one kind of thing that corresponds to all these different kinds of occurrences of indices.

Now it might be argued that none of this matters because scarcely anybody actually does situation semantics anymore anyway; even in the HPSG community, the Pollard-Sag situation-semantics-inspired CONTENT values have mostly been superseded, following (Richter, 2000; Richter and Sailer, 1999) by CONTENT values that are essentially feature-structural encodings of terms of higher-order logic (usually Ty2). Unfortunately, this does not make it any easier to say just what exactly indices are supposed to be. If CONTENT values are just encodings of Ty2 formulas, this means that the index of a name is a constant; the index of a deictic pronoun is a free variable; the index of a pronoun whose antecedent is a QNP is a bound variable; the index of a direct-object reflexive where the subject is a name is the same constant as that corresponding to the subject; and the index of a QNP is ... what?

This last case is especially problematic, because in a logical translation of of a sentence containing a QNP, where the variable corresponding to the QNP occurs (and indeed, *whether* it occurs) depends on essentially stylistic decisions about the form of the translation. For example, consider the sentence *every boy runs*. For precisely which of the imaginable ways of translating this sentence into Ty2 is the feature-structure encoding of that translation the “real” CONTENT value of *every boy runs*? Is it $\text{every}'(\text{boy}')(\text{run}')$, which contains no variable occurrences at all? Or is it perhaps the familiar first-order reduction $\forall x(\text{boy}'(x) \rightarrow \text{run}'(x))$? Or, as Quine might have had it, $\lambda x \top = \lambda x(\text{boy}'(x) \rightarrow \text{walk}'(x))$? Given the conventional wisdom that lambda-terms are dispensable—only the denotation in a model, which is invariant under term equivalence, is supposed to matter—it

shouldn't make any difference. But for the HPSG binding theory to work, it is crucial that the indices, whatever they are, be located at the ends of precisely the right paths in the feature structures, so that we can know exactly where to look for the substructures on whose token-(non)identity the entire BT hinges.

Of course one can require that the (feature-structure encodings of) lambda terms corresponding to CONTENT values of nondummy NPs and nonpredicative PPs be written in precisely the right style to guarantee that the right subterm always shows up in the encoding at the end of such-and-such a path; but then it seems evident that there is no natural class of empirical phenomena that BT is constraining; instead one is essentially deciding in advance which kinds of sentences one wants ruled in (or out) by BT and then cooking the representations to ensure that those cases are covered. To put it another way, in the world of real phenomena, there is no such thing as the index of a noun phrase. (We believe this to be a noncontroversial assertion.) So what are we to make of a theory whose predictive power is based on whether or not two given NPs in an utterance have the same index? We will revisit this question in connection with Principle B in section 4.

3 Refining the theory of English r-pronouns

3.1 English r-pronouns according to Pollard and Xue

It is well known (Zribi-Hertz 1989, Baker 1995) that, at least in certain literary (especially British) registers, referred to here as Lit./Brit., and under suitable pragmatic conditions, even locally o-commanded reflexives may fail to be coindexed with a local o-commander. The examples in (6) illustrate this point:

- (6) a. (...) his_i wife was equally incredulous of her innocence and suspected himself_i, the pastor, to be the cause of her distress, (...)
(Zribi-Hertz 1989: (37))
- b. Philip_i was supposed to be fooling (...), because Desiree (...) had undoubtedly explained to them the precise nature of her relationship with himself_i.
(Zribi-Hertz 1989: (43b))
- c. But Rupert_i was not unduly worried about Peter's opinion of himself_i.
(Zribi-Hertz 1989: (46b))

In (6a), the object reflexive refers not to the pastor's wife, as Principle A predicts, but rather to the pastor, whose narrative point of view is being reflected. In (6b) and (6c), both of which have a reflexive prepositional object within a PNP, the reflexive refers not to the possessor as predicted by Principle A, but rather someone else who is somehow prominent in the discourse, perhaps the topic or perhaps an individual whose mental state is being described. Zribi-Hertz assumed that in such examples, logophoricity was the operative factor. Baker argued for a different notion of contrastive intensification involving reference to discourse prominent entities. To account for such facts, Pollard and Xue (1998, 2001) proposed the theory of English r-pronouns given in (7):

- (7) A theory of English r-pronouns (Pollard and Xue 1998, 2001):
- a. Principle R:
Every r-pronoun is either
 - i. coindexed with a local o-commander, or
 - ii. interpreted according to certain pragmatic constraints involving logophoricity, contrastiveness, or discourse prominence.
 - b. Principle A as per (2) (colloquial American English only)

On this account, Principle A is simply absent from Lit./Brit. Instead, it is assumed that a weaker constraint, Principle R, applicable to English in general, requires that any r-pronoun *either* be coindexed with a local o-commander *or* pragmatically constrained (inclusive disjunction)

3.2 English reflexives according to Runner and Kaiser

Pollard and Xue continued to assume that Principle A as stated in (2) applies to colloquial American English. But as Runner and Kaiser—and other recent work that they cite, much of it based on carefully controlled experimental investigations—have shown, this cannot be right, because of examples like the ones in (8) where the prepositional object inside a picture NP is not coindexed with the possessive determiner:

- (8) Counterexamples to Pollard/Sag form of Principle A
- a. Ebenezer_i saw Jacob's_j picture of himself_{i/j}.
(Runner and Kaiser 2005: (7))
 - b. Manray burned Mary's photo of himself.
(Runner and Kaiser 2005: (28))

Accordingly, Runner and Kaiser propose to amend Principle A as shown in (9):

- (9) Revised Principle A (Runner and Kaiser 2005):
Every locally o-commanded reflexive is coindexed with a local d-commander.

Pace Pollard and Sag (and Pollard and Xue), on this account specifiers (including possessive determiners) are not classified as valents (syntactic arguments). Instead, Specifiers and (at least some) adjuncts, as well as valents, are classified as **dependents** (in the sense of Bouma et al. 2001— except that for them, specifiers are not subsumed under dependents). Local d-command is then defined in the same way with respect to dependents as local o-command is with respect to valents. In particular, possessive determiners locally d-command (but crucially do not locally o-command) PP complements of the head noun. With these revisions, the reflexives in (8) become BT-exempt.

3.3 A synthesis

Thus Pollard and Xue on the one hand, and Runner and Kaiser on the other, both propose to relax Pollard and Sag's Principle A, in quite different ways, to cover somewhat different sets of facts. Where the two proposals appear to come into competition is in accounting for examples like (6)b,c. However, if we accept that Runner and Kaiser are right about possessive determiners not being arguments, then Principle R can be allowed to stand as stated in (7); it would still cover examples like (6a), but examples like (6)b,c would now be accounted for by Runner and Kaiser's formulation of Principle A. This tentative conclusion is summarized in (10):

(10) Tentative synthetic theory of r-pronouns for English

- a. As per Runner/Kaiser, possessive determiners are dependents (not valents), and locally d-command (not o-command) PP complements of the N.
- b. Principle R (7a)
- c. Principle A as per Runner/Kaiser (9) (colloquial American English only)

3.4 A simpler theory?

Can the account in (10) be simplified? Pollard and Xue misassessed the facts, perpetuating the widespread but erroneous belief that examples like (8) were limited to a certain elevated register. But could it be that in reality English r-pronouns, even locally o-commanded ones like the one in (6a), are *never* obligatorily coindexed with a local d-commander, not even in colloquial American English? If so, we could simplify our account to the one in (11):

(11) A possible simpler theory of English r-pronouns

- a. As per Runner/Kaiser, possessive determiners are dependents (not valents), and locally d-command (not o-command) PP complements of the N.
- b. Principle R':
Every r-pronoun is either
 - i. coindexed with a local d-commander, or
 - ii. interpreted according to certain pragmatic constraints involving logophoricity, contrastiveness, or discourse prominence.

In this theory, Principle R is modified to make reference to local d-command rather than local o-command, and Principle A is dropped altogether. Choosing between this account and the one in (10) comes down to the factual question of whether examples like (6)a are really restricted to a certain register or not. It seems that what is required in order to answer this question is to apply Runner and Kaiser's experimental methodology to a wider range of sentence types, along the lines of (12):

- (12) More facts to assess with varying head types
- a. Manray burned Mary's photo of himself.
(noun)
 - b. Manray burned Mary's tasteless critique of himself.
(deverbal noun)
 - c. Manray was outraged at Mary's tasteless criticism of himself.
(deverbal noun)
 - d. Manray was outraged at Mary's tasteless criticizing of himself.
(nominal gerund)
 - e. Manray was outraged at Mary's tastelessly criticizing himself.
(verbal gerund with possessive subject)
 - f. Manray was outraged at Mary tastelessly criticizing himself.
(verbal gerund with accusative subject)
 - g. Manray was outraged that Mary tastelessly criticized himself.
(finite verb)

Note that these examples form a cline from purely nominal to purely verbal constructions, with constructions headed by various kinds of deverbal nouns and gerunds occupying the middle ground. Is there a point on the cline beyond which the intended readings are no longer available in colloquial American English? The simpler theory hypothesized in (11) predicts that there is not.

3.5 Problems with predicative NPs

Before leaving the topic of English r-pronouns, we briefly consider some seldom-discussed cases that we think merit more careful investigation. It is rare for a head to have both a subject and a specifier (in the limited HPSG sense), which perhaps is one reason why in GB theory subjects were always subsumed under the notion of specifier. But, at least in HPSG terms, one environment where both can occur with the same head is in predicative NPs with possessive determiners, illustrated the examples in (13) and (14):

- (13) Predicative NPs with possessive determiners I: absolutes
(Pollard and Sag 1994: (56))
- a. With [Kim and Sandy]_i [each other's]_i closest confidants, it will be good for them to have a chance to do some travelling together.
 - b. *With Kim_i his_i greatest admirer, it's obvious that he isn't going to win any popularity contests.

(14) Predicative NPs with possessive determiners II: complements
(Pollard and Sag 1994: (57)-(58))

- a. [Kim and Sandy]_i are [each other's]_i greatest admirers.
- b. *[Kim and Sandy]_i are their_i greatest admirers.
(cf. [Kim and Sandy]_i met their_i greatest admirers.
- c. We_j consider [Yeltsin and Gorbachev]_i to be [each other's]_{i/*j} greatest potential allies.
- d. We_j consider Gorbachev_i to be our_j/*his_i greatest admirer.
(cf. We_j consider Gorbachev_i to have met his_i greatest admirer.

In fact examples like these were Pollard and Sag's motivation for proposing that in cases where both a subject and a specifier occur, the subject locally o-commands the specifier. But now, since we are no longer treating possessive determiners as valents, we must modify this assumption to the form in (15):

(15) If a predicative NP has a possessive determiner, it is locally d-commanded by the NP's subject.

But then what are we to make of examples like the ones in (16)?

- (16) a. John considers Mary the polar opposite of himself.
b. Mary treats her friends as mere extensions of herself.

In each of these examples, the reflexive PP complement is locally o-commanded by the unexpressed subject of the predicative NP, which in turn is controlled by (and therefore coindexed with) the matrix object. Thus the synthetic theory in (10) wrongly excludes these examples, while the simpler theory in (11) allows them.

But then, how are we to explain (14)c? In that example, according to the simpler theory (11), the reciprocal possessive determiner should be able to have as its antecedent *either* the locally d-commanding unexpressed subject of the predicative NP, which is controlled by the matrix object *Yeltsin and Gorbachev*, *or* the matrix subject *we*, which denotes the individuals whose point of view or mental state is being reported. But Pollard and Sag judged this second ("logophoric") interpretation to be unavailable. Was that judgment simply mistaken? The structurally similar example (17) suggests that it may well have been:

(17) [Kim and Sandy]_i consider loyalty to be [each other's]_i most admirable trait.

If this example is acceptable, then it provides further support for the simpler theory.

4 Principle B Reconsidered

We turn now to the seemingly simpler question of how to formulate Principle B. For ease of reference, Runner and Kaiser's formulation of Principle B is given in (18) together with Chomsky's (1986) formulation. Both are paraphrased slightly in order to emphasize their essential similarity:

(18) Two formulations of Principle B

- a. (Runner/Kaiser 2005) No p-pronoun is coindexed with any of its local d-commanders. [**local** = being a codependent of the same head as the p-pronoun].
- b. (Chomsky 1986) No governed pronominal is coindexed with any of its local c-commanders. [**local** = being in an argument position in the governing category of the pronominal.]

Of course these two formulations employ different notions of command and locality; but otherwise they are strikingly similar, especially in assuming that the non-dummy NPs in a sentence actually have things called indices whose identity or lack thereof can be sensibly theorized about.

But of course NP utterances do not come with indices stamped on them that we can check for identity or nonidentity. So then how do we tell whether some version or other of Principle B is making correct predictions? To get a clearer understanding of just what is at issue here, consider the sentences in (19):

(19) Examples typically taken as confirming Principle B

- a. John saw him.
- b. He saw him.
- c. Every man saw him.
- d. Who did Mary say saw him?
- e. John tried to see him.

These are all examples that would typically be taken as confirming evidence for the correctness of Principle B, but in each case the *reason* for taking them as confirming Principle B is different, as shown in (20):

(20) Reasons for taking the examples in (19) as confirming evidence for Principle B: the sentence cannot mean

- a. $\text{see}'(j, j)$
- b. $\text{see}'(x, x)$
- c. $\text{every}'(\text{man}', \lambda x.\text{see}'(x, x))$
- d. $\lambda p.\text{some}'(\text{person}', \lambda x.p = \text{say}'(x, \text{see}'(x, x)))$
- e. $\text{try}'(j, \lambda x.\text{see}'(x, x))$

Here we are representing the impossible interpretations by fairly standard lambda terms in some form or other of intensional or hyperintensional logic. If we try to articulate just what it is that is being disallowed, in terms of the syntactic forms of the lambda terms, it seems different in each case, as shown (21):

(21) What is being ruled out, in terms of the form of the logic translation

- a. A pronoun cannot be translated by the same constant as a locally commanding name.

- b. A pronoun cannot be translated by the same variable as a locally commanding pronoun.
- c. A pronoun cannot be translated by the same variable as the one bound by the λ operator corresponding to the scope of the semantic determiner that translates the determiner of a locally commanding NP.
- d. A pronoun cannot be translated by the same variable as the one bound by the λ operator corresponding to an unbounded dependency/A-bar movement if the trace/tail of the corresponding chain locally commands the pronoun.
- e. A pronoun cannot be translated by the same variable as the one bound by the λ operator arising from the translation of a locally commanding “unexpressed subject” (PRO).

The point here is just that if we think of Principle B as a constraint on the syntactic forms of logical expressions that denote the interpretations in question, it does not seem to be expressing a coherent empirical generalization.

One step toward making the five cases above look more alike is to make the assumption in (22):

(22) Assumption about translation of names:

Unless they are de-accented, utterances of names are translated by fresh parameters (variables that cannot be bound).

If this is right, then it seems we may have a chance of dispensing with the notion of index altogether in favor of a constraint on how logical translations are assigned to linguistic expressions, as sketched in (23):

(23) Toward a reformulation of Principle B as a constraint on logical translation:

No p-pronoun is translated by the same variable as any of its local d-commanders.

This of course is of course only a programmatic proposal, not a theory. In order to make sense of it, we must take a lot for granted, including the following

(24) Some presuppositions of (23)

- a. The translation of every nondummy NP, including unrealized ones (such as trace and PRO) consists (at least in part) of an occurrence of a logical variable in an argument position of some atomic formula in the translation (or in some elementary predication in the sense of minimal recursion semantics⁵ (MRS, Copestake et al. (in press))).

⁵In terms of MRS (or an algorithm for assembly of “unplugged” lambda terms roughly analogous to it), nonidentity of variables per Principle B could be implemented by introducing **variable constraints** of the form $x \neq y$, where x, y, \dots are not true variables but “prevariables” to be replaced by variables subject to the variable constraints at the resolution (or meaning assembly) stage. Such prevariables could be thought of as bearing the same relation to variables over individuals (or individual concepts) that MRS handles bear to variables over propositions.

- b. Some elements (QNPs, PRO, trace, etc.) also contribute a (λ) operator that binds the contributed variable.
- c. Assembly of the pieces of logical syntax contributed by the words of an utterance into a single term denoting the interpretation of the utterance is in some sense independent of, or subsequent to, the satisfaction of Principle B.

The working out the technical details of such a programmatic proposal is evidently a long-term enterprise, and not one that we are eager to undertake, but this is what seems to be required if we want logical variables to take up the explanatory burden that has been borne by indices.

5 Conclusion

In the preceding remarks, we have suggested some revisions of English BT in response to certain classes of data that were problematic for previous formulations. The simpler formulation (11) of Principle A seems to provide a straightforward picture of the connection between reflexivity, argument structure, and logophoricity, though further experimental investigation is needed in order to confirm or disconfirm its predictions. However, this account shares with its many predecessors the uncomfortable positing of indices as theoretical primitives. Eliminating indices in favor of logical variables presents itself as an obvious alternative, but even in the seemingly simpler case of Principle B, where the complicating factors of reflexivity and logophoricity are absent, the technical obstacles to be overcome seem daunting.

Is there a better way? We suspect that there might be. Though the interpretation of pronouns as variables has a venerable pedigree (traceable at least as far back as Montague's PTQ), another possibility, proposed by Pollard and Sag 1983 and Pollard 1984, is that referentially dependent elements (including p-pronouns, r-pronouns, and traces) denote identity functions, and that predicates combine with them not by function application but rather by composition. Unfortunately Pollard and Sag did not pursue this line of investigation, because identity functions, function composition, and their ilk are not comfortably accommodated in the impoverished type theory upon which HPSG is based.⁶ However, these same ideas were taken up independently by Jacobson starting in the early 1990s and developed into a highly promising research program (see e.g. Jacobson 1999, 2000, and other works cited there).

Until recently, Jacobson's line of investigation focused on aspects of anaphora orthogonal to BT. However, in recent unpublished work (Jacobson ms.), she has set her sights on Principle B. We did not learn of this in time to make a proper assessment here, but the gist of it is roughly as follows. As in Jacobson's other work, expressions containing an "unbound" pronoun are of a different syntactic type than ones that do not; roughly, they have an implicative (i.e. functional or

⁶What is missing is the exponential type-constructor, which gives rise to functional types.

conditional) type, where the antecedent of the conditional is NP (corresponding to the pronoun). What is new is that the functions interpreting such expressions have their domains restricted so as to induce the effect of Principle B, e.g. the interpretation of *Bush praises him* is a function whose domain excludes Bush.

The adjustment of the syntactic type is necessitated by Jacobson's adherence to a principle of strict compositionality in which each syntactic type corresponds to a unique semantic type. This differs from the Pollard and Sag (1983) approach, in which a single syntactic type can correspond to a range of semantic types: a "basic" one for expressions that do not contain any unbound referentially dependent subexpressions, as well as implicative types with that basic type as the final consequent for ones that do. Another difference is that on the Pollard and Sag (1983) approach, the antecedent types corresponding to unbound referentially dependent subexpressions are semantic types, not syntactic ones, reflecting the apparent lack of syntactic connectivity.⁷ A third difference is that Jacobson employs only one implicative type constructor (aside from the categorial left and right slashes), whereas Pollard (1984) used different "binding features" (the forerunners of HPSG's nonlocal features) for p-pronouns, r-pronouns, relative pronouns, and interrogative pronouns; on a type-logical recasting of HPSG along the lines of Pollard 2004, these binding features would correspond to different flavors of implication, and "cobound" pronouns to type-logical shifts of the form $A \Rightarrow (A \Rightarrow B) \vdash A \Rightarrow B$. We leave the consideration of these and related issues for future exploration.

References

- Carl Lee Baker. 1995. Contrast, discourse prominence, and intensification, with special reference to locally free relatives in British English. *Language* 71:63-101.
- Gosse Bouma, Robert Malouf, and Ivan Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19:1-65.
- António Branco, Francisco Costa, and Manfred Sailer, eds. 2005. *Binding Theory and Invariants in Anaphoric Relations: Binding Theory Workshop Notes*. Technical report TR-05-11, Departamento de Informática, Universidade de Lisboa.
- Noam Chomsky. 1986. *Knowledge of Language*. Praeger, New York.
- Ann Copestake, Daniel Flickinger, Carl Pollard, and Ivan Sag. In press. Minimal recursion semantics: an introduction. To appear in *Research on Language and Computation*.
- Martin Everaert. 1991. Contextual distinction of the anaphor/pronominal distinction. In J. Koster and E. Reuland, eds.
- Karin Golde. 1999. *Binding Theory and Beyond: An Investigation into the English Pronominal System*. Ph.D. dissertation, The Ohio State University.
- Lars Hellan. 1991. Containment and connectedness anaphors. In J. Koster and E. Reuland, eds.

⁷Except in the case of traces, where the syntactic category of the trace has to be "remembered" in order to enforce syntactic connectivity in unbounded dependency constructions.

- Pauline Jacobson. 1999. Toward a variable-free semantics. *Linguistics and Philosophy* 22:117-184.
- Pauline Jacobson. 2000. Paycheck pronouns, Bach-Peters sentences, and variable-free semantics. *Natural Language Semantics* 8(2):77-155.
- Pauline Jacobson. Ms. Direct compositionality and variable-free semantics: the case of “Principle B” effects. Manuscript, Brown University.
- Jan Koster and Eric Reuland. *Long-Distance Anaphora*. Cambridge University Press, New York.
- Carl Pollard. 1984. *Generalized Context-Free Grammars, Head Grammars, and Natural Language*. Ph.D. dissertation, Stanford University.
- Carl Pollard. 2004. Type-logical HPSG. In G. Jäger, P. Monachesi, G. Penn, and S. Wintner, eds., *Proceedings of Formal Grammar 2004*. Nancy, France, European Summer School in Language, Logic, and Information, pp. 107-124.
- Carl Pollard and Ivan A. Sag. 1992. Anaphors in English and the scope of the binding theory. *Linguistic Inquiry* 23:261-303.
- Carl Pollard and Ivan A. Sag. 1983. Reflexives and reciprocals in English: an alternative to the binding theory. In M. Barlow, D. Flickinger, and M. Wescoat, eds., *Proceedings of the Second West Coast Conference on Formal Linguistics*, Department of Linguistics, Stanford University, pp. 189-203.
- Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press and CSLI Publications, Chicago and Stanford.
- Carl Pollard and Ping Xue. 1998. Chinese reflexive *ziji*: syntactic reflexives vs. nonsyntactic reflexives. *Journal of East Asian Linguistics* 7:287-318.
- Carl Pollard and Ping Xue. 2001. Syntactic and nonsyntactic constraints on long-distance reflexives. In P. Cole, G. Hermon, and C.T. James Huang, eds., *Long-Distance Reflexives*. Syntax and Semantics v. 33, 317-342. Academic Press, San Diego.
- Tanya Reinhart and Eric Reuland. 1991. Anaphors and logophors: an argument structure perspective. In J. Koster and E. Reuland, eds.
- Tanya Reinhart and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657-720.
- Frank Richter. 2000. *A Mathematical Formalism for Linguistic Theories with an Application to Head-Driven Phrase Structure Grammar*. Ph.D. dissertation, Universität Tübingen.
- Frank Richter and Manfred Sailer. 1999. LF conditions on expressions of Ty2: an HPSG analysis of negative concord in Polish. In R. Borsley and A. Przepiórkowski, eds., *Slavic in HPSG*. CSLI Publications, Stanford, pp. 247-282.
- Jeffrey T. Runner and Elsi Kaiser. 2005. Binding in picture noun phrases: implications for binding theory. In A. Branco et al., eds., pp. 53-58.
- Jeffrey T. Runner, Rachel S. Sussman, and Michael K. Tannenhaus. 2002. Logophors in possessed picture noun phrases. *Proceedings of the 21st West Coast Conference on Formal Linguistics*, 401-414.
- Peter Sells. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18:445-479.

Anne Zribi-Hertz. 1989. Anaphora binding and narrative point of view: English reflexive pronouns in sentences and discourse. *Language* 65:695-727.