Linearity-Based Reference Restrictions in Hindi

The problem   As illustrated in (1), Hindi disallows coreference between an R-expression and a pronoun even if the latter does not c-command the former. This restriction is surprising because, due to the lack of c-command, no binding principle is violated. Thus, nothing should block accidental coreference between *sher* 'lion' and *us-ke* 'his' (Heim & Kratzer 1998), contrary to fact.

(1)   Raam-ne [us-ke$_1$ baccõ-ko]   sher$_{2/*1}$ dikhaayaa
       Ram-ERG he-GEN children-DAT lion   show
       'Ram showed his$_1$ children a lion$_{2/*1}$.'

The structure becomes grammatical if the positions of the R-expression and the pronoun are reversed, as in (2), or if the R-expression is moved over the pronoun, as in (3).

(2)   Raam-ne sher$_{1/2}$ [us-ke$_1$ baccõ-ko]   *t* dikhaayaa
       Ram-ERG lion   he-GEN children-DAT   show
       'Ram showed its$_1$ children a lion$_{1/2}$.'

(3)   Raam-ne [sher-ke$_1$ baccõ-ko] us-ko$_{1/2}$ dikhaayaa
       Ram-ERG lion lion-GEN children-DAT him show
       'Ram showed a lion's$_1$ children to him$_{1/2}$.'

For ease of reference, I will call these disjointness effects *extended disjoint reference (EDR) effects* as they obtain in configurations that the standard binding principles do not apply to. EDR effects are systematic in that they arise in an entirely parallel manner between subjects and objects as well. They likewise occur if the lower element is a proper name rather than a referential common noun.

The role of locality   EDR effects are not plausibly reduced to standard Principle C violations. While Principle C effects are global, EDR effects are strictly local. In particular, if the two elements are separated by an island boundary, EDR effects disappear while standard Principle C effects remain:

(4)   a.   EDR effects between an R-expression and a pronoun disappear if either of the two is properly embedded in an island that the other is outside of.
       b.   Principle C effects persist in this configuration.

This generalization can be demonstrated by means of possessors. Possessors in Hindi demonstrably disallow subextraction and are hence islands. Interestingly, if either of the two nominals is embedded in a possessor, no EDR effects occur. (5) contains a proper name embedded within the possessor of the direct object. The indirect object *us-ke baccõ-ko* contains a pronoun. Coreference is possible, in stark contrast to (1). What sets the two cases apart is the fact that the R-expression is inside an island in (5) but not in (1).

(5)   mãĩ us-ke$_1$ baccõ-ko   [[Raam-ke$_1$ bacpan-kii]   tasviirẽ] dikhaanaa caahtaa hũ
       I   he-GEN children-DAT Ram-GEN childhood-GEN pictures show.INF   want   be.1SG
       'I want to show his$_1$ children pictures of Ram's$_1$ childhood.'

Crucially, Principle C effects differ from EDR effects in that they are not affected by islands. The sentence in (6) is parallel to the one in (5) except for the fact that the pronoun c-commands the R-expression rather than being embedded in a c-commanding noun phrase. Coreference is ruled out. That there are environments where Principle C effects persist while EDR effects do not strongly suggests that the latter cannot be reduced to the former. The same pattern is observed for other islands, e.g., subject clauses or gerunds.

(6)   *mãĩ us-ko$_1$ [[Raam-ke$_1$ bacpan-kii]   tasviirẽ] dikhaanaa caahtaa hũ
       I   he-DAT Ram-GEN childhood-GEN pictures show.INF   want   be.1SG
       'I want to show him$_1$ pictures of Ram's$_1$ childhood.'

EDR effects also disappear if the pronoun, rather than the R-expression, is embedded in a possessor. Contrast (7a), where coreference is ruled out, with (7b), where no disjoint reference effects obtains, due to the possessor

island the pronoun is embedded in.

(7) a. [us-kii$_1$ mãã]   Raam-ko$_{*1/2}$ pasand kartii hai
    he-GEN mother Ram-ACC   love   do   be.3SG
    'His$_1$ mother loves Ram$_{*1/2}$.'

    b. [[us-kii$_1$ mãã]   -kii behen] Raam-ko$_{1/2}$ pasand kartii hai
    he-GEN mother -GEN sister   Ram-ACC   love   do   be.3SG
    'His$_1$ mother's sister likes Ram$_{1/2}$.'

This set of data is captured by the descriptive generalization in (4).

Proposal: Linearity-based reference restrictions   To account for this constellation of facts I propose that the relevant principle operates on linear precedence rather than hierarchical structures. This explains without further ado why c-command does not seem to play a role and why movement of the R-expression (in (3)) allows coreference. In particular, I suggest that there is restriction on linear strings demanding linear precedence relations to correspond to the obviative hierarchy in (8). This principle is stated more precisely in (9), where '>' designates linear precedence and 'LIN' the linearization algorithm. (9) is a well-formedness requirement on the output of the linearization procedure.

(8) *Obviative Hierarchy* (Safir 2004)
    R-expression $\gg$ pronoun $\gg$ anaphor

(9) *Obviative Alignment*
    Given a syntactic structure Σ, such that $\text{LIN}(\Sigma) = \langle \ldots > x_i > \ldots > x_j > \ldots \rangle$:
    If x_i and x_j are coindexed then $x_i \gg x_j$.

To see how (9) applies to a concrete example, consider the sentence in (7a). Linearization of the hierarchical syntactic structure yields the string in (10a). I assume, contra much literature on linearization, that referential indices are part of the linearized representation. In (10a), obviative alignment (9) is violated because the R-expression *Raam* follows the coreferential pronoun *us-kii* but is more obviative than it. Coreference is thus excluded. If the linear position between the two elements is reversed (as in (2) or (3)) obviative alignment is adhered to and coreference becomes possible. This accounts for the contrast between (1) vs. (2) and (3).

(10) a. $\langle$ us-kii$_1$ > mãã > Raam-ko$_1$ > pasand > kartii > hai $\rangle$
    b. (i)   us-kii$_1$ > Raam-ki$_1$
       (ii) us-kii$_1$ $\not\gg$ Raam-ki$_1$ $\rightsquigarrow$ *violates (9)*

The locality of EDR effects   The fact that coreference across island boundaries is allowed follows under a concept of *Cyclic Linearization*, according to which islands are linearization domains (Uriagereka 1999, Chomsky 2000, Stjepanović & Takahashi 2001, Johnson 2004, Bošković 2007). The idea underlying these proposals is that linearization renders a syntactic structure simplex, hence preventing subsequent extraction out of it. This concept of islandhood, coupled with (9) as a filter on the output of individual instances of the linearization algorithm, derives the grammaticality of (5) and (7b) as follows. Since the possessor is an island, it is linearized separately from the rest of the clause. The output of linearizing the island is given in (11a). The second cycle linearizes the entire clause. Importantly, the possessor, having already undergone linearization, is a syntactically simplex element at this point. It's individual parts are not subject to renewed linearization. The output is provided in (11b). Both outputs in (11) conform to (9) and coreference is hence possible. This derives the fact that EDR effects only show up if no island boundary intervenes between the two elements.

(11) *Linearization domains of (5)*
    a. LIN$_1$ = $\langle$ Raam-ke$_1$ > bacpan-kii $\rangle$
    b. LIN$_2$ = $\langle$ mãĩ > us-ko$_1$ > [LIN$_1$] > tasviirẽ > dikhaanaa > caahtaa > hũ $\rangle$