

Excursus: Bresnan (1976) on the A-over-A Principle

Fabian Heck & Gereon Müller

Institut für Linguistik
Universität Leipzig

WiSe 2009/2010

www.uni-leipzig.de/~muellerg

In view of the evidence that the A-over-A Principle is too strong, there are only two strategies that can be pursued.

- The first possibility is to give up the constraint, and replace it with some other constraint(s). This is the position taken by Ross (1967).
- The second possibility is to revise the constraint in such a way that the counter-examples lose their force (while, ideally, supporting evidence can be maintained). A revision of the A-over-A Principle is argued for by Bresnan (1976) (on the basis of earlier work in Chomsky (1973)).

Chomsky (1973) on the A-over-A Principle

(1) **A-over-A Principle** (Chomsky's (1973) version):

If a transformation applies to a structure of the form

$$[\alpha \dots [A \dots] \dots]$$

where α is a cyclic node, then it must be so interpreted as to apply to the maximal phrase of the type A.

Chomsky (1973, 235) remarks that this version of the A-over-A Principle “does not establish an absolute prohibition against transformations that extract a phrase of type A from a more inclusive phrase of type A. Rather, it states that if a transformational rule is nonspecific with respect to the configuration defined, it will be interpreted in such a way as to satisfy the condition.”

This understanding of the A-over-A Principle already captures fundamental aspects of Bresnan's revision. However, it does not necessarily follow from the formulation in (1), at least not as long as it is not clarified what exactly it means for a transformation to “apply to a phrase of type A”. As Chomsky notes, “alternatively, one might interpret the A-over-A constraint as legislating against any rule that extracts a phrase of type A from a more inclusive phrase A,” which is the standard understanding. He concludes that “the former interpretation [...] is perhaps more natural”, and therefore “tentatively” adopts it in the remainder of the study.

Bresnan (1976): Background

Bresnan (1976) develops an elaborate theory of transformations; on this basis, she suggests a revision of the A-over-A Principle that evades the counter-evidence given before. The basic idea is that the A-over-A Principle should not be sensitive to category information; rather, it is sensitive to the nature of the transformation involved. The central claim is that the item that can be subjected to a given transformation should be as inclusive as possible, independently of category labels (whereas an even more inclusive item that cannot be subjected to the same transformation because it does not fit its structural description can be ignored, even if it shares the same category label). Consider the following revision of the A-over-A Principle taken from Bresnan (1976, 14). (Bresnan takes this to be essentially a more formal implementation of Chomsky's (1973) basic idea.)

Bresnan's (1976) A-over-A Principle

- (2) **A-over-A Principle** (Bresnan's revision):
No transformation T can apply to a structure ϕ under a proper analysis π unless π is a maximal proper analysis of ϕ for T .

This version of the principle relies on the notion of “maximal proper analysis”; it is defined in (3).

- (3) **Maximal proper analysis:**
 π is a maximal proper analysis of ϕ for $T = \langle \mathcal{C}, \mathcal{M} \rangle$ iff π assigns maximal values to all target predicates in S .

Key notions involved here: A transformation T is an ordered pair $\langle S, \mathcal{M} \rangle$, where S is a **structural condition** consisting of a number of terms, and \mathcal{M} is a **transformational mapping** affecting these terms. The *wh*-movement transformation, e.g., can be encoded as the $\langle S, \mathcal{M} \rangle$ pair in (4) (I tacitly adapt parts of Bresnan's original formulation to more current terminology).

(4) **Wh-movement transformation:**

$$\begin{array}{ccccccc}
 [_{CP} & Q & - & W_1 & - & [_{XP} & wh & - & W_2 &] & - & W_3 &] \\
 & 1 & & 2 & & & \underbrace{3 \quad 4} & & & & & 5 & \Rightarrow \\
 & \underbrace{3 \quad 4} & & 2 & & & \emptyset & & & & & 5 & \\
 \end{array}$$

Here, Q represents the specifier of an interrogative CP. The transformation moves the *wh*-phrase from its in situ position to Q , as an instance of substitution, and carries out deletion in the base position of the *wh*-phrase.

Variables, Predicates, Proper Analysis

Variables (W_1 – W_3) vs. **predicates**: Variables stand for any sequence of terms whereas predicates encode restrictions for the transformation, defining \mathcal{S} . Two types of predicates:

- **Target predicates** are the items that the transformation applies to (the item that is moved, in the case at hand).
- **Context predicates** are the non-target predicates of a structural condition (i.e., items that must be present in a phrase structure for the transformation to be able to apply, but which are not directly affected by the operation).

- (5) a. A **proper analysis** of a syntactic structure ϕ for a transformation $T = \langle \mathcal{S}, \mathcal{M} \rangle$ is a factorization $\langle \phi_1, \dots, \phi_n \rangle$ of ϕ (simplifying, a complete partition of the whole initial phrase marker into subtrees) of which \mathcal{S} is true and upon which \mathcal{M} is defined.
- b. A **maximal proper analysis** as in (3) is a proper analysis of a phrase marker that “assigns maximal values to all target predicates in \mathcal{S} ”; and the value of a target predicate is maximal if it represents the most inclusive category (i.e., the one that dominates all others) among those that satisfy the requirements imposed by \mathcal{S} on the transformation.

Thus: The revised A-over-A Principle in (2) states that a transformation can only apply to a given structure if the size of the target predicate is maximal vis-à-vis the demands imposed by the structural condition \mathcal{S} of the transformation.

Categories vs. Structural Conditions on Transformations

Before going through an example illustrating the working of this version of the A-over-A Principle, it is worth emphasizing that there is a significant shift of perspective: Category information is at the heart of the original A-over-A Principle, but it does not play any role whatsoever here. Two (or more) categories compete for movement not if they have identical labels, but if they both satisfy requirements imposed by the structural condition of a transformation on the form of the target predicate. Bresnan (1976, 21ff) highlights this new property of the A-over-A Principle by focussing on “cross-categorial” transformations, i.e., transformations that are not category-specific (like, arguably, movement to subject position in English is). *Wh*-movement as in (4) is just such a transformation; see (6).

- (6)
- a. What book did you read ?
 - b. How long is it ?
 - c. How quickly did you read it ?
 - d. How much did it cost ?

A Problem for the Old A-over-A Principle

Consider now the data in (7) (from Bresnan (1976, 22), with (7-de) added), which therefore pose a problem for the original version of the A-over-A Principle (in that they show that the constraint is too weak).

- (7)
- a. How many feet tall does the girl stand ?
 - b. *How many feet does the girl stand tall ?
 - c. *How many does the girl stand feet tall ?
 - d. *How does the girl stand many feet tall ?
 - e. How many feet are in a yard ?

Note:

Parallel examples from German show a slightly different pattern, with the construction that is analogous to (7-b) emerging as the preferred option; see Heck (2008).

- (8)
- a. Wieviele Zentimeter ist sie groß ?
how many centimeters is she tall
 - b. ?Wieviele Zentimeter groß ist sie ?
how many centimeters tall is she

The present logic might lead one to hypothesize that in German, an analysis of the pre-movement string is available according to which *wieviele Zentimeter* ('how many centimeters') emerges as the most inclusive item.

Bresnan's Analysis

Bresnan assumes that the pre-movement structure of *how many feet tall* in (8-abc) is roughly as in (9).

(9) Q the girl [_{VP} stands [_{AP} [_{NP} [_{QP} [_{DP} how] [Q many]] [N feet]] [A tall]]]

Possible proper analyses for *wh*-movement:

- a factorization in which the DP *how* is the moved
- a factorization in which the QP *how many* is the moved item
- a factorization in which the NP *how many feet* is the moved item
- a factorization in which the AP *how many feet tall* is the moved item.

Here the A-over-A Principle becomes relevant and blocks (8-bcd), leaving (8-a) as the sole remaining option. Only (8-a) involves a **maximal proper analysis**.

Note:

An even more inclusive category would be the VP that minimally dominates AP in (9). However, movement of the VP *stands how many feet tall* does not meet the requirements of *S* of (4) according to which the *wh*-item must be leftmost in the moved phrase. As noted by Bresnan (1976), there is nothing inherently wrong with moving a category like the NP in (9). If there is no more inclusive XP that could act as a *wh*-phrase, as is the case in (7-e), movement of such an NP is possible.

This general reasoning is quite independent of the exact structure attributed to categories in the adjectival and nominal domain, and of the labelling of the categories (although details of the analysis may differ under other assumptions). Suppose, e.g., that we were to assume (as one of several further options for analysis that have been proposed) (10) instead of (9) as the structure that feeds *wh*-movement.

(10) Q the girl [_{VP} stands [_{AP} [_{DP} [_{DegP} [_{Deg} how] [_{QP} many]]] [_{D'} [_D Ø] [_{NP} feet]]] [_A tall]]]

In this case, Bresnan's A-over-A Principle in (2) would imply that AP movement is possible whereas DP and DegP movement are not; movement of *how* alone would be excluded by the requirement that it has to be an XP that undergoes the movement.

Exactly the same kind of analysis can be given for the examples in (11) (without commitments to structural analysis).

- (11)
- a. You have a very intelligent sister
 - b. How intelligent a sister do you have ?
 - c. *How intelligent do you have a sister ?

Whatever the internal structure of the nominal projection here is supposed to look like, it is clear that movement of the more inclusive *wh*-category in (11-b) respects Bresnan's A-over-A Principle whereas movement of the less inclusive *wh*-category in (11-c) violates it.

However, not all of the evidence that can be brought forward against the original A-over-A Principle (showing that it is too weak) can be derived under Bresnan's reformulation. For instance, the ban on preposition stranding in German (see (12)) does not follow from (2) either (at least not unless further assumptions are made).

(12)

- a. Sie spielt [PP₁ mit [DP₂ dem grünen Auto]]
she plays with the green car
- b. [PP₁ Mit [DP₂ welchem Auto]] spielt sie t₁ ?
with which car plays she
- c. *[DP₂ Welchem Auto] spielt sie [PP₁ mit t₂] ?
which car plays she with
- d. [PP₁ Mit [DP₂ dem grünen Auto]] spielt sie t₁
with the green car plays she
- e. *[DP₂ Diesem Auto] spielt sie [PP₁ mit t₂]
this car plays she with

Evidence in Support of the original A-over-A Principle

What about the evidence *in support* of the original A-over-A Principle (see (13), (14))? The ill-formed examples here involve illicit *wh*-movement (or relativization) of a *wh*-item (or a relative pronoun) where the more inclusive category has the same category feature but cannot act as a *wh*-item (or relative item) itself. These data can therefore not be derived from Bresnan's A-over-A Principle.

(13)

- a. [DP₁ My letter to [DP₂ a friend in Italy]] got lost
- b. *[DP₂ Who] did [DP₁ my letter to t₂] get lost ?
- c. [DP₁ Which letter to [DP₂ a friend in Italy]] got lost?
- d. *John is the friend [DP₂ who] C [DP₁ my letter to t₂] got lost
- e. This is the letter [DP₁ which] t₁ got lost

(14)

- a. John heard [DP₁ a rumour that you had read [DP₂ this book]]
- b. *[DP₂ What] did John hear [DP₁ a rumour that you had read t₂] ?
- c. [DP₁ Which rumour that you had read [DP₂ this book]] did John hear ?
- d. *This is a book [DP₂ which] John heard [DP₁ a rumour that you had read t₂]
- e. This a rumour [DP₁ which] John heard t₁

Evidence Against the original A-over-A Principle

Finally, what about the evidence that proved that the original A-over-A Principle is too strong?

(15)

- a. [DP₂ Who] would you approve of [DP₁ my seeing t₂] ?
- b. [DP₂ Which author] did you read [DP₁ a book about t₂] ?

First, (15) (where movement of a *wh*-DP takes place from a more inclusive non-*wh*-DP, and the result is well formed) ceases to be a problem for exactly the reason that (13), (14) cannot be derived anymore as ungrammatical: Categorical information as such is irrelevant, and movement of a lower *wh*-item cannot be blocked by a higher non-*wh*-item. If one compares the ungrammatical examples in, say, (13) with the grammatical examples in, e.g., (15), it is indeed hard to see how any version of the A-over-A Principle could make the right distinctions. The evident difference is one between ill-formed extraction from subject and well-formed extraction from object, but this is a domain that the A-over-A Principle has nothing to say about, in any version. Hence, in light of the wellformedness of data like those in (15), the fact that the revised A-over-A Principle is not capable of excluding data like those in (13) should not be viewed as a shortcoming; the data suggest that some other locality constraint that can distinguish between subjects and objects will account for the difference.

More Evidence Against the original A-over-A Principle

Second, the data in (16) have CP topicalization from a CP. Assuming (as seems natural) that the lower CP has an abstract feature [top] that is required by the structural description \mathcal{S} of an appropriately defined topicalization transformation, and that the higher CP does not have such a feature, the two items do not interact, and Bresnan's A-over-A Principle makes the right predictions; again, categorial identity emerges as irrelevant.

(16)

- a. John wouldn't say [_{CP₁} that Mary thinks [_{CP₂} that Bill is nice]]
- b. [_{CP₂} That Bill is nice] John wouldn't say [_{CP₁} that Mary thinks t₂]
- c. Fritz hat behauptet [_{CP₁} Maria würde denken [_{CP₂} dass er nett ist]]
Fritz_{nom} has claimed Maria_{nom} would think that he nice is
- d. [_{CP₂} Dass er nett ist] hat Fritz behauptet [_{CP₁} würde Maria denken t₂]

Yet More Evidence Against the original A-over-A Principle

Third, the same analysis can be given for the VP-over-VP contexts that permit topicalization of the lower VP in (17) and (18): The lower VP that undergoes the movement differs from the higher VP (or VPs) in that it bears the [top] feature required by the structural condition of the topicalization transformation.

- (17)
- a. Fritz hat [_{VP₁} [_{VP₂} zu arbeiten] versucht]
Fritz_{nom} has to work tried
 - b. [_{VP₂} Zu arbeiten] hat Fritz [_{VP₁} t₂ versucht]
to work has Fritz_{nom} tried
 - c. [_{VP₁} [_{VP₂} Zu arbeiten] versucht] hat Fritz t₁
to work tried has Fritz_{nom}
- (18)
- a. [_{VP₂} Zu arbeiten] denke₄ ich nicht [_{VP₀} t₄ [_{CP} dass er [_{VP₁} t₂ versucht]
to work think I now that he tried
hat]]
has
 - b. [_{VP₁} [_{VP₂} Zu arbeiten] versucht] denke₄ ich nicht [_{VP₀} t₄ [_{CP} dass er t₁
to work tried think I not that he
hat]]
has

Is the A-over-A Principle a Good Constraint?

The version of the A-over-A Principle in (2) meets the demands of a good constraint (as does the original version): Abstracting away from the technical aspects of Bresnan's analysis, it is simple and general, and it is not complex.

The formulation in (2) is not the definitive version that Bresnan (1976) ends up with. Rather, eventually she suggests a “**Relativized A-over-A Principle**” in which the notion of a “maximal proper analysis” is replaced with the weaker notion of an “**r-maximal proper analysis**” (“weaker” in the sense that there are more r-maximal proper analyses than there are maximal analyses). The basic idea is that the size of target predicates is not maximized as such; target predicates are only maximized according to a given choice of context predicates. The motivation for this change is based on issues like the increased options of affecting target predicates in deletion constructions (like VP ellipsis).

Incidentally, as explicitly noted by Bresnan (1976, 16), the version of the A-over-A Principle in (2) already is a “relativized” constraint – relativized with respect to possible items that can be affected by transformations (rather than being rigidly defined in terms of categorial labels).

Towards the F-over-F Principle

From the perspective of current linguistic theory, Bresnan's (1976) revision of the original A-over-A Principle looks very modern. In a theory where movement transformations involve designated features, her approach can straightforwardly be reinterpreted as requiring movement of the closest item that bears a feature matching the attracting feature on the head of the landing site. From this perspective, (2) amounts to stating that if there is a head X with a movement-inducing feature \mathcal{F} , of all the items Y_1, \dots, Y_n that are in the c-command domain of X and bear a matching feature \mathcal{F} , only the most inclusive Y_i (i.e., the Y_i that dominates all other Y_j 's) can be moved to the specifier of X . This reformulation of Bresnan's (2) is a constraint that has in fact more recently been proposed by a number of researchers, as an alternative to the original A-over-A Principle (and, it may be noted, without reference to Bresnan (1976) (or, for that matter, Chomsky (1973)) throughout). Thus, assuming that certain designated features are responsible for triggering various movement operations ([*wh*] for *wh*-movement, [*top*] for topicalization, etc.), it is clear that these features are not (necessarily) categorial (they are "cross-categorial", in Bresnan's terms).

The F-over-F Principle

On this basis, it has been proposed that the original A-over-A Principle should be revised as in (19); here and henceforth, I will refer to this minimality-based revision as the **F-over-F Principle**. See Takano (1994), Koizumi (1995), Fukui (1997), Kitahara (1997), Müller (1998), Sauerland (1999), Fitzpatrick (2002), Vicente (2007), and Heck (2008), among others.

(19) **F-over-F Principle:**

In a structure $\alpha_{[*F*]} \dots [\beta_{[F]} \dots [\gamma_{[F]} \dots] \dots] \dots$, movement to $[*F*]$ can only affect the category bearing the $[F]$ feature that is closer to $[*F*]$.

(19) is very similar to (2). There is a difference with respect to the empirical domains that are covered by the analyses employing these two kinds of constraints, though. Bresnan only concerns herself with initial ambiguities in rule application that arise from the fact that more than one category can in principle represent the moved item for one and the same feature. Essentially, all her data center around the question of whether **pied piping** is possible (and the formulation of (4) suggests that leftmost items can pied-pipe other material whereas non-leftmost items cannot). In contrast, in the approaches that rely on (some version of) (19), the empirical domain typically does not involve pied piping but rather two separate *wh*-phrases, one of which dominates the other in the pre-movement structure.

F-over-F and Remnant Movement: Topicalization in German

Empirical evidence in support of a constraint like the F-over-F Principle comes from the consideration of a restriction on the movement of remnant categories (i.e., categories from which extraction has taken place). Among other things, the constraint blocks certain illicit instances of remnant scrambling in languages like German and Japanese (see Takano (1994), Koizumi (1995), Kitahara (1997), Müller (1998), Sauerland (1999)). Consider first some data from German. (20-ab) show that a restructuring infinitive from which scrambling has taken place (i.e., a remnant infinitive) can be topicalized.

- (20) a. [_α t₁ Zu lesen]₃ hat [_{DP} das Buch]₁ keiner t₃ versucht
to read has the book no-one tried
- b. [_α t₁ t₂ Zu reparieren]₃ hat der Frank dem Matthias₁ den Drucker₂ t₃
to fix has the Frank_{nom} the Matthias_{dat} the printer_{acc}
versprochen
promised

Next, (21-ab) illustrate that the same remnant infinitive from which scrambling has taken place cannot undergo scrambling itself (scrambling of a non-remnant infinitive is possible, though).

- (21) a. *dass [_α t₁ zu lesen]₃ [_{DP} das Buch]₁ keiner t₃ versucht hat
that to read the book_{acc} no-one tried has
- b. *dass der Frank [_α t₁ t₂ zu reparieren]₃ dem Matthias₁ den Drucker₂ t₃
that the Frank_{nom} to fix the Matthias_{dat} the printer_{acc}
versprochen hat
promised has

F-over-F and Remnant Movement: Scrambling in Japanese

The same restriction is illustrated on the basis of data from Japanese in (22). In (22-a), a complement CP has undergone scrambling in front of the subject; (22-b) shows that such scrambling is impossible if DP scrambling has taken place from the complement CP prior to CP fronting.

- a. $[_{CP}$ Mary-ga $[_{DP}$ sono hon-o $]]_1$ yonda-to $]]_2$ Bill-ga $[_{CP}$ John-ga t₂
Mary_{nom} that book_{acc} read-COMP Bill_{nom} John_{nom}
itta-to $]]$ omotteiru (koto)
said-COMP think (fact)
- b. * $[_{CP}$ Mary-ga t₁ yonda-to $]]_2$ $[_{DP}$ sono hon-o $]]_1$ John-ga t₂ itta (koto)
Mary_{nom} read-COMP that book_{acc} John_{nom} said (fact)

This anti-identity restriction on remnant movement (which is referred to as an **Unambiguous Domination** requirement in Müller (1998)) is not confined to scrambling. The same effect shows up with *wh*-movement in English: Extraction of a *wh*-phrase from a *wh*-phrase makes subsequent *wh*-movement impossible; the effect is much stronger than one would expect if only a typical *wh*-island violation with *wh*-argument extraction were involved.

- (23) a. $*[_{DP} \text{ Which book about } t_1]_2 \text{ don't you know } [_{CP} \text{ who}_1 \text{ to read } t_2] ?$
b. $*[_{DP} \text{ Which picture of } t_1]_2 \text{ do you wonder } [_{CP} \text{ who}_1 \text{ she likes } t_2] ?$

In the ungrammatical cases, there is a stage of the derivation where a movement-inducing feature (like $[\ast\Sigma\ast]$ for scrambling, and $[\ast wh\ast]$ for *wh*-movement) on a target head could in principle attract either the more inclusive category or another category dominated by the latter (both bearing, by assumption, a matching feature $[\Sigma]$ or $[wh]$). The F-over-F Principle then demands movement of the higher category (β in (19)) first (which is thus not yet a remnant category at this point because extraction from it has not yet taken place), and subsequent movement of a category (α in (19)) from within β to a lower position will invariably be excluded by the Strict Cycle Condition, which (among other things) excludes cases of lowering.

I will return to the status of the F-over-F Principle below. For the time being, we may leave it at that, and turn to other predecessors of current locality constraints, viz., the island constraints proposed by Ross (1967), at least partly in reaction to Chomsky's A-over-A Principle.

References

- Bresnan, Joan (1976): On the Form and Functioning of Transformations, *Linguistic Inquiry* 7, 3–40.
- Cable, Seth (2007): The Grammar of Q. PhD thesis, MIT, Cambridge, Mass.
- Chomsky, Noam (1973): Conditions on Transformations. In: S. Anderson & P. Kiparsky, eds., *A Festschrift for Morris Halle*. Academic Press, New York, pp. 232–286.
- Fitzpatrick, Justin (2002): On Minimalist Approaches to the Locality of Movement, *Linguistic Inquiry* 33, 443–463.
- Fukui, Naoki (1997): Attract and the A-over-A Principle, *UCI Working Papers in Linguistics* 3, 51–67.
- Heck, Fabian (2004): A Theory of Pied Piping. PhD thesis, Universität Tübingen.
- Heck, Fabian (2008): *On Pied-Piping. Wh-Movement and Beyond*. Vol. 98 of *Studies in Generative Grammar*, Mouton de Gruyter, Berlin.
- Kitahara, Hisatsugu (1997): *Elementary Operations and Optimal Derivations*. MIT Press, Cambridge, Mass.
- Koizumi, Masatoshi (1995): Phrase Structure in Minimalist Syntax. PhD thesis, MIT, Cambridge, Mass.
- Müller, Gereon (1998): *Incomplete Category Fronting*. Kluwer, Dordrecht.
- Ross, John (1967): Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, Mass.
- Sauerland, Uli (1999): Erasability and Interpretation, *Syntax* 3, 161–188.
- Stechow, Arnim von & Wolfgang Sternefeld (1988): *Bausteine syntaktischen Wissens*. Westdeutscher Verlag, Opladen.
- Takano, Yuji (1994): Unbound Traces and Indeterminacy of Derivation. In: M. Nakamura, ed., *Current Topics in English and Japanese*. Hituzi Syobo, Tokyo, pp. 229–253.
- Vicente, Luis (2007): The Syntax of Heads and Phrases. PhD thesis, LOT: Netherlands Graduate School of Linguistics, Universiteit Leiden.
- Wurmbrand, Susanne (2001): *Infinitives. Restructuring and Clause Structure*. Mouton de Gruyter, Berlin.