Across-the-Board Extraction in Minimalist Grammars

Gregory M. Kobele
Institut für deutsche Sprache und Linguistik
Humboldt-Universität zu Berlin
kobele@rz.hu-berlin.de

Abstract

Minimalist grammars cannot provide adequate descriptions of constructions in which a single filler saturates two mutually independent gaps, as is commonly analyzed to be the case in parasitic gap constructions and other across-the-board extraction phenomena. In this paper, I show how a simple addition to the minimalist grammar formalism allows for a unified treatment of control and parasitic gap phenomena, and can be restricted in such a way as to account for across-the-board exceptions to the coordinate structure constraint. In the context of standard constraints on movement, the weak generative capacity of the formalism remains unaffected.

1 Introduction

Minimalist grammars (MGs) (Stabler, 1997) are a mildly context-sensitive grammar formalism (Michaelis, 2001), which provide a rigorous foundation for some of the main ideas of the minimalist program (Chomsky, 1995). There are two basic structure building operations, binary **merge** and unary **move**.

In typical analyses of linguistic phenomena, operator–variable chains (in the sense of a quantifier and its bound variable; $\forall x.\phi[x]$) are analyzed in terms of first merger of the operator into a position where its variable is ultimately to appear (resulting in something like $\phi[\forall]$), and then moving the operator into its scope-taking position, leaving a bound 'trace' in the moved-from position (resulting in the desired $\forall x.\phi[x]$).[1]

Now, although minimalist grammars can capture naturally various kinds of non-local dependencies in this way, something needs to be added to the system to allow it to account for apparent non-resource-sensitive behaviour. Pursuing the logical formula metaphor introduced above, MGs can define only (closed) *linear* formulae, where each variable is bound by a distinct quantifier, and each quantifier binds exactly one variable. However, the phenomena of control and parasitic gaps both involve a single filler being associated with multiple gaps—in other words, the 'chains' here are tree-structured (see fig.1).[2]

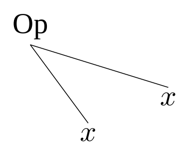

Figure 1: The filler-gap dependencies exemplified by parasitic gaps

In this paper, we show how slash-feature percolation, as adapted to MGs by Kobele (2007), allows for a straight-forward implementation of

[1]This non-compositional description of the intermediate derivational steps is for the imagination only. Compositional semantics for this kind of analysis are easy to provide (see Kobele (2006)), as it is after all just the familiar Cooper-storage (Cooper, 1983) writ funny.

[2]For the sake of perspicuity, I am ignoring multiple movements of the same subexpression, as occurs in the analysis of passivization followed by wh-movement, as in a sentence like *Who did John think* `t` *was kissed* `t`. The distinguishing characteristic which sets the control/parasitic gap type phenomena apart from this kind of multiple movement is that in this case each movement is to a c-commanding position in the derived structure—such movement chains have the shape of strings, whereas those of the parasitic gap variety are trees.

Sag's (1983) analysis of parasitic gap phenomena in the minimalist framework, while preserving the weak generative capacity of the formal system. This analysis extends immediately to control. Other cases of such non-resource-sensitive phenomena, such as well-known exceptions to the coordinate structure constraint, fall out as well, although a (weak generative capacity preserving) extension to the minimalist grammar type-system is needed to account for some of the familiar restrictions on such movements.

2 Slash-feature percolation and MGs

In the minimalist tradition, where long distance dependencies are mediated via movement, across-the-board extraction out of a conjunct as in 1 is sometimes thought to be derived from an intermediate structure of the form below:

1. Who did John meet and Susan kiss?

$$[_S\ John\ meet\ who]\ and\ [_S\ Susan\ kiss\ who]$$

In order for this kind of analysis to work, some mechanism must be in place to ensure the identity of both moving elements—identity of *derived structure*, not merely of category (as suggested by example 2).

2. *Which bank did John rob or Susan walk along?

Crucially, this mechanism is not reducible to ellipsis in this framework, as it must allow a single resource (the trigger for movement residing in the COMP position) to meet the requirements of multiple expressions (the features on each of the wh-words)—ellipsis is not standardly assumed to have this character.

A simple way around this problem is to introduce the ATB-moved element *after* conjoining the two clauses together. To implement this idea, we adopt the mechanism of slash-feature percolation, as adapted to MGs by Kobele (2007). (Slash-feature introductions are represented in the below as traces.)

$$[_S\ John\ meet\ \mathtt{t}]\ and\ [_S\ Susan\ kiss\ \mathtt{t}]$$

The change required to Kobele's system is to allow *identical slash-features* to be *unified*, instead of crashing the derivation of an expression. (This is simply Sag's (1983) GPSG analysis adapted to this framework.) An MG expression can be represented as a tuple of categorized structures (each element of this tuple corresponds to a moving treelet). In order to make the link with LCFRSs, a finite upper bound needs to be placed on the length of such a tuple. Stabler proposes that no two treelets may have the same first feature (this amounts to a strict version of Chomsky's Shortest Move Constraint). Kobele maintains this assumption in his enriched MG system as well, forcing missing (i.e. 'slashed') expressions to behave in the same way as real ones. Our proposal builds on the fact that slashed expressions, unlike moving expressions, have no internal structure. Thus, there arises no computational problem in comparing two slash-features—it is an atomic operation. Specifically, we claim that in order to avoid shortest move violations, identical slash-features may be unified with one another. A specific instance of the **merge** operation is given in figure 2. An expression of the form $(\text{John meet} : \mathtt{S}), \langle \mathtt{d\ \text{-}k\ \text{-}q}, \mathtt{\text{-}w} \rangle$ indicates that it is selectable as a tensed sentence ($\mathtt{S}$), and that it is missing an element of type $\mathtt{d\ \text{-}k\ \text{-}q\ \text{-}w}$ (a +wh noun phrase), but that it has satisfied the first three dependencies ($\mathtt{d\ \text{-}k\ \text{-}q}$) of this expression. As the expression it is merged with in this figure is missing the very same type of element, they are identified in the result. This contrasts with the situation in which one (or both) of the wh-phrases is already present, as in figure 3. In this case, the resulting expression has two subexpressions (the slashed expression $\langle \mathtt{d\ \text{-}k\ \text{-}q}, \mathtt{\text{-}w} \rangle$ and the wh-phrase $(\text{who}, \mathtt{\text{-}w})$) with the same active first feature ($\mathtt{\text{-}w}$), violating the shortest move constraint.

3 Control and Parasitic Gaps

With this slight relaxation of resource sensitivity with respect to hypotheses, we are able to account for control (3) and parasitic gap constructions (4) in terms of across the board movement.

3. John wanted to kiss Susan.

4. Who did John want to kiss before meeting?

Essentially, slash-feature unification gives us the ability to have limited *sideward* movement in the sense of Nunes (2004). The analysis here of ATB

$$\frac{(\text{and Susan kiss} : \texttt{=S S}), \langle \texttt{d -k -q}, \texttt{-w}\rangle \qquad (\text{John meet} : \texttt{S}), \langle \texttt{d -k -q}, \texttt{-wh}\rangle}{(\text{John meet and Susan kiss} : \texttt{S}), \langle \texttt{d -k -q}, \texttt{-w}\rangle}$$

Figure 2: Unification of hypotheses avoids crash

$$\frac{(\text{and Susan kiss} : \texttt{=S S}), \langle \texttt{d -k -q}, \texttt{-w}\rangle \qquad (\text{John meet} : \texttt{S}), (\text{who}, \texttt{-w})}{(\text{John meet and Susan kiss} : \texttt{S}), \langle \texttt{d -k -q}, \texttt{-w}\rangle, (\text{who}, \texttt{-w})}$$

Figure 3: An SMC violation

movement and of parasitic gaps can be seen as a (clear and precise) variant of Nunes'.[3]

3.1 Control

The treatment of control as (a form of) movement agrees in spirit with recent developments in the minimalist tradition (Hornstein, 2001; Kobele, 2006), but its unification with *ATB* movement forces us to make the base position of the controller not c-command the base position of the controllee. Instead, the controller must raise to a position which c-commands the controllee, as sketched in figure 4.

$$[_{VP}\ [_{V'}\ \texttt{t}\ ]\ [_{S} \ldots \texttt{t} \ldots]\ ]$$

Figure 4: Control as ATB Movement

In the context of minimalist grammars, this movement is naturally identified with movement for case (to an object agreement position—AgrOP). The treatment of control as a form of movement obviates the need for the empty category PRO, and thus of mysterious indices relating controllee and controller. Instead, 'PRO' is simply a trace in a theta-position, coindexation is replaced by chain formation, and the effects of the 'control module' need to be enforced via standard constraints on movement.

A fragment for English which implements these ideas is given as figure 5. The fragment is the same as the one given in Kobele (2006) (which is to say that this treatment of control is broadly compatible with other standard analyses), except that base positions of sentential complements and arguments in obligatory control verbs have been altered so as to conform to the anti-c-command condition imposed by this analysis of control.[4]

This grammar gets both subject and object control constructions, as in 5 and 6 below.

5. John promised Mary to shave every barber.

6. John persuaded Mary to shave every barber.

The object control case is perhaps the most surprising, as the object (in 6, *Mary*) is supposed to move outside of the VP, and yet clearly follows the verb. The basic idea of the analysis of such cases is that movement for case does indeed put the object to the left of the verb, but that subsequent head movement of the verb (broadly following Chomsky's (1957) affix hopping analysis of the English auxiliary system) remedies the situation. The choice of controller (whether subject or object) in the sentences 5 and 6 above is determined by whether the sentential complement is merged before or after the base position of the subject. If before, there is no subject slash-feature to be unified with the slash-feature in the sentential complement, and thus subject control is impossible. If after, then the SMC ensures that the object must already have checked its case (as the more

[3] The difference is, of course, that here the phonological content of the ATB-moved expression is not present at its base positions, whereas in Nunes' system, it is. The present treatment of control in terms of ATB movement cannot analyze purported cases of 'backward control' (Polinsky and Potsdam, 2006) as such.

[4] Due to space limitations, only intuitions about relevant aspects of this grammar will be attempted to be conveyed here. The interested reader is invited to consult Kobele (2006) for definitions and examples. The fragment deals with raising, expletive *it*, control, passivization, and quantifier scope, in a way that neatly derives the ban on super-raising, and the tensed-clause-boundedness of QR.

that::=S s		
will::=perf +k +q S	have::=en perf	be::=ing prog
-s::=>perf +k +q S	-en::=>prog en	-ing::=>v ing
-ed::=>perf +k +q S	ϵ::=>prog perf	ϵ::=>v prog
to::=perf s	be::=pass v	-en::=>V pass
	ϵ::=>V +k =d +q v	
arrive::=d v	devour::=d V	
	shave::=d V	
seem::=s v	expect::=s V	expect::=d =s v
	want::=s V	want::=d =s v
	hope::=S V	hope::=d =s v
persuade::=d =s V		
promise::=s =d V	promise::=d +k =d =s +q v	
ϵ::=>v =z v	it::z -k	
George::d -k -q	the::=n d -k -q	ointment::n
John::d -k -q	every::=n d -k -q	abbot::n
Mary::d -k -q	some::=n d -k -q	barber::n

Figure 5: A grammar for English A movement (slightly modified from Kobele (2006))

recently merged subject has case requirements of its own), and thus slash-features of the sentential complement have only the slash-features of the matrix subject to unify with.

3.2 TAG Approaches to Control

Both Kroch and Joshi (1985) as well as the XTAG project (2001) make use of the empty category PRO to mediate control relations between argument positions. In the XTAG project use is made of equations in feature structures, which allow for lexical determination of binding relationships. Restrictions on the relative positions of controller and controllee, as well as on the realization of PRO is governed by formalism external constraints on elementary trees.

The Multi-dominance TAG system of Chen-Main (2006) seems able to implement a PRO-less theory of control, along the lines proposed in Hornstein (2001). Her account of node contraction as being constrained by derivational locality might provide a principled (formalism internal) account of the configurational relation between controller and controllee.

3.3 Parasitic Gaps

In summary, this extension to MGs gives us a formal system in which fillers can be associated with multiple gaps in certain circumstances. Evaluation of the linguistic applicability of such a formalism needs to be done with respect to the kinds of analyses that it makes available. We continue to assume the analysis of English given as figure 5. We have sketched the implementation of obligatory control present in this fragment above. Parasitic gaps (of the form in 4) require an analysis of gerundival adjuncts, which we will treat here as vP adjuncts (i.e. they appear after the logical subject, if any, is introduced), which we will analyze in terms of the **adjoin** operation formalized in Frey and Gärtner (2002).[5] A gerundival adjunct is headed by a

[5] Another option is to explicitly control the direction of merger (via merge-left (x=) and merge-right (=x) features), and then to mediate adjunction via an empty lexical item ϵ::=>v =pg v, where prepositions heading gerundival adjuncts are assigned the type =g pg. The basic problem is that the verb needs to remain accessible for future affixal operations, and that the gerundival adjunct needs to appear to the right of the vP. With the ‘standard’ antisymmetric treatment of the linear order of merged elements (first-merged object to the right, all others to the left) embedded in classical MGs, this is not so easily done, as in order for the verb to remain accessible to future head movement operations, it must

prepositional element which selects a gerundive clause (before::=g ≈v), and a gerundive clause is simply what you get if instead of merging a tense marker (like will::=perf +k +q S or to::=perf s) you merge *-ing* (-ing::=>perf g). At the vP level (and at the perfP level), a clause will have a DP waiting to check its case (-k) and scope (-q) features (the logical subject, if the clause is in the active voice, and the logical object, otherwise). In order to avoid a violation of the SMC, both this DP in the vP and the DP in the prepositional gerundive adjunct must be slashed expressions (of the form ⟨d, -k -q⟩) which are identified, resulting in a control configuration. A parasitic gap configuration arises when the object of the prepositional gerundive clause and of the vP are slashed as well (of the form ⟨d -k -q, -w⟩). Note that this analysis accounts for the well-known fact that A movement (passive or raising) does not license parasitic gaps (as in 7): the object in the prepositional genitive has had its case and scope features checked, and thus can only survive as a slashed expression if it is moving again (say, to check a *wh* feature). For this reason it will not be unifiable with an A moving expression, which is looking next to check its case feature. Of course, a sentence like 7 can be made grammatical by passivizing the gerundive adjunct clause (as in 8), thereby taking it out of the parasitic gap construction type.

7. *Susan was kissed before meeting.

8. Susan was kissed before being met.

There are many properties of the parasitic gap construction that this analysis does not account for (see the collection Culicover and Postal (2001)), but it does capture some interesting properties in a simple way, without changing the weak generative capacity of the system (the proof of this is essentially the same as the one given in Michaelis (2001)—it is a consequence of the fact that the number of possible hypotheses are upper-bounded by the number of distinct licensee (-x) features in the grammar (due to the SMC)).

The ability to license parasitic gaps, as mentioned briefly above, has often been used as a diagnostic to distinguish between A and A-bar movement types. The analysis of parasitic gaps given here suggests that the fact that only A-bar (i.e. *wh*) movements license parasitic gaps can be explained in purely configurational terms, i.e. without positing some occult connection between movement types and the parasitic gap phenomenon. This has, as far as this analysis is on the right track, serious ramifications for analyses of phenomena (such as scrambling), which have used parasitic gap licensing properties to argue that these phenomena involve a particular dependency type. This perspective on parasitic gaps also makes the fact that in certain languages A movements can license parasitic gaps unsurprising.

3.4 Parasitic Gaps in TAGs

Kroch and Joshi (1985) account for parasitic gap constructions as in 4 by taking as elementary trees biclausal structures. They note that they make a grammatical distinction between sentences like 4 on the one hand, and sentences like 9 on the other, where the parasitic gap is embedded in a subordinate clause inside of the prepositional gerund introducing it.

9. What did John devour without bothering to shave?

They are forced into this position by their treatment of gerunds as non-syntactically derived words. If one retreated to a post-syntactic view of morphology, whereby the terminal items in elementary trees were thought of not as concrete words but rather as abstract lexemes, whose ultimate realization were determined in part by their syntactic context, then as long as the 'gerundivizing' feature were above a clausal adjunction site, and the verb below it, sentences such as 9 would be generable. Thus, this aspect of Kroch and Joshi's analysis seems almost an accidental property.

A more fundamental difference between the Kroch and Joshi proposal and the one presented here lies in the number of gaps which can be associated with a single overt operator. As their strategy is to extend the size of elementary trees so that all gaps and binders are contained within them, they are forced to the position that there is a fixed upper bound on the number of parasitic gaps that can be dependent on one operator. In the present theory, parasitic gaps can be recursively embedded inside others, and thus there is predicted to be no principled upper bound on the number of gaps

be merged first, and thus the merger of the gerundival adjunct places it erroneously to the left of the vP.

$$\begin{array}{lll} \text{what::}\mathtt{d}\ \mathtt{-k}\ \mathtt{-q}\ \mathtt{-w} & \text{before::}\mathtt{=g} \approx\mathtt{v} & \text{-ing::}\mathtt{=>perf}\ \mathtt{g} \\ \epsilon\text{::}\mathtt{=>S}\ \mathtt{+w}\ \mathtt{c} & \text{without::}\mathtt{=g} \approx\mathtt{v} & \end{array}$$

Figure 6: Extending the grammar in fig. 5 to get parasitic gap constructions

a single filler can fill. A sentence distinguishing these two proposals is given below as 11 (where Kroch and Joshi's original proposal of two is taken as the cut-off point).

10. What did Mary take without paying for?

11. What did Mary intend to return after taking without paying for?

Unfortunately, neither sentence 9 nor sentence 11 are the 'clear cases' (Chomsky, 1956) that one should base linguistic theories on. As such, their ultimate grammatical status will have to wait an independent validation of one or the other syntactic theory.

4 Coordinate Structures

Although slash-feature identification seems to capture the basic effect of ATB movement (as noted by Sag), the prototypical ATB construction (as in 1) involves extraction out of a conjunction, and is constrained in ways that we currently do not account for (in particular, extraction must be out of *both* conjuncts). To implement these constraints, we need a way to block movement out of an expression, and to ensure that the slash-features of both conjuncts are identical. As a first step, we build this in to the category system in the following way. First, we add a diacritic on category features ($\mathtt{c}^*$) which permits them to be selected only if they have no moving elements (this can be implemented as a restriction on the domain of the **merge** operation, and allows for the simple statement of a certain kind of island constraint). To be able to ensure that the slash-features of two different selected items are identical, we need to more drastically revise the minimalist category system. We want to assign the following type to *and*:

$$\mathtt{=S}^{\alpha}\ \mathtt{=S}^{\alpha}\ \mathtt{S}^*$$

The interpretation of the superscripted material is as the slash-features of the selected expression, and the fact that both superscripts are identical on both selection features indicates that both selected expressions must have the same slash-features.[6] At least intuitively, a link may be drawn between this extension of the minimalist grammar type-system, and the addition of local-constraints (Kroch and Joshi, 1985) to tree-adjoining grammars.

Chen-Main (2006) further develops the system introduced in Sarkar and Joshi (1996) to deal with conjunction in TAGs. So long as the contractible nodes in any given derived tree are bounded in advance, it seems as though a strategy like the one pursued here could be extended to her system. The elementary tree for *and* would have equations on each of its two substitution nodes stating that, *after* substitution and subsequent internal *and*-tree internal contractions, the remaining contractible nodes in each subtree are identical. The remaining contractible nodes in each subtree would need to be identified with an identical partner in the other subtree, so as to ensure that later 'movement' be truly across the board.

5 Conclusion

Slash-feature percolation is straightforward to add to the minimalist grammar formalism, and amounts to relaxing the requirement that an expression must be derivationally present before it can begin satisfying dependencies. This paper has tried to show that adding slash-feature percolation to MGs allows for interesting and revealing analyses of what I have been calling across-the-board extraction phenomena, analyses which coexist well with other analyses of different grammatical phenomena. Indeed, it seems not implausible that slash-feature percolation (which, in MGs, acts upon the regular derivation trees) is simply addable to TAGs as well (but on the level of the derived tree), and can make similar analytical options available as it does for MGs, and CFGs.

[6]This is like passing the same stack of indices to two non-terminals in an indexed grammar. The essential difference is that our 'stacks' are bounded in size (due to the SMC). This fact makes this enrichment of the type system weakly innocuous. Note that this very same constraint is stateable in GPSG, without any increase in generative capacity.

References

Chen-Main, Joan. 2006. *On the generation and linearization of multi-dominance structures*. Ph.D. thesis, Johns Hopkins.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124.

Chomsky, Noam. 1957. *Syntactic Structures*. Mouton, The Hague.

Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press, Cambridge, Massachusetts.

Cooper, Robin. 1983. *Quantification and Syntactic Theory*. D. Reidel, Dordrecht.

Culicover, Peter W. and Paul M. Postal, editors. 2001. *Parasitic Gaps*, volume 35 of *Current Studies in Linguistics*. MIT Press.

Frey, Werner and Hans-Martin Gärtner. 2002. Scrambling and adjunction in minimalist grammars. In Jäger, Gerhard, Paola Monachesi, Gerald Penn, and Shuly Wintner, editors, *Proceedings of Formal Grammar 2002*, pages 41–52.

Hornstein, Norbert. 2001. *Move! A Minimalist Theory of Construal*. Blackwell.

Kobele, Gregory M. 2006. *Generating Copies: An investigation into structural identity in language and grammar*. Ph.D. thesis, University of California, Los Angeles.

Kobele, Gregory M. 2007. A formal foundation for A and A-bar movement in the minimalist program. In Kracht, Marcus, Gerald Penn, and Edward P. Stabler, editors, *Mathematics of Language 10*. UCLA.

Kroch, Anthony S. and Aravind K. Joshi. 1985. The linguistic relevance of tree adjoining grammar. Technical Report MS-CIS-85-16, Department of Computer & Information Science, University of Pennsylvania.

Michaelis, Jens. 2001. *On Formal Properties of Minimalist Grammars*. Ph.D. thesis, Universität Potsdam.

Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*, volume 43 of *Linguistic Inquiry Monographs*. MIT Press, Cambridge, Massachusetts.

Polinsky, Maria and Eric Potsdam. 2006. Expanding the scope of control and raising. *Syntax*, 9(2):171–192, August.

Sag, Ivan A. 1983. On parasitic gaps. *Linguistics and Philosophy*, 6:35–45.

Sarkar, Anoop and Aravind K. Joshi. 1996. Coordination in tree adjoining grammars: Formalization and implementation. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, pages 610–615, Copenhagen, Denmark.

Stabler, Edward P. 1997. Derivational minimalism. In Retoré, Christian, editor, *Logical Aspects of Computational Linguistics*, volume 1328 of *Lecture Notes in Computer Science*, pages 68–95. Springer-Verlag, Berlin.

XTAG Research Group. 2001. A lexicalized tree adjoining grammar for English. Technical Report IRCS-01-03, IRCS, University of Pennsylvania.