Remarks on Movement
David Adger
December, 2018

Syntactic dependencies can be very local, as when a head combines with a complement, or when a head-complement structure takes a specifier. If this is all there was to syntax, the techniques of early structuralist approaches, such as those of Bloomfield (1926), appropriately updated and formalised as a phrase structure grammar, would be sufficient to provide a theoretical understanding of the grammatical component of human language. Each element would bear a single set of relations to each other element. A head could select its complement (or specifier), assign case to it, determine its thematic role, agree with it, etc. Each element would have its unique place in structure.

But syntactic dependencies can also be non-local. A single element, say a noun phrase (or DP in more updated parlance), may relate to other elements locally, and non-locally. Take the passive construction in English for example:

(1) Lilly was bitten.

The noun phrase *Lilly* relates to the tensed auxiliary *was* in terms of agreement, and to the T functional head more generally in terms of case. It is, at least arguably, local in terms of hierarchy and linear order, to these elements of structure (it is the specifier of T and the auxiliary *was* is structurally in T). However, *Lilly* also relates to the participle *bitten*, in that the root of that participle, the verb *bite*, determines the thematic role that *Lilly* plays; that verb *bite* also syntactically selects *Lilly*. But *Lilly* is not local to *bite*. It is neither the complement nor the specifier of that verb and it is linearly separated from it by *was*. This kind of non-locality can be quite extreme, as in the following case.

(2) Lilly is believed to have been bitten.

Here, a fair amount of structure separates *Lilly* and *bite,* both linearly and hierarchically.

Similarly, in (3), the wh-phrase *which cat*, is syntactically local to the auxiliary *did*, but it receives its thematic interpretation from the verb *bit*, which is in an entirely separate finite clause.

(3) Which cat did Anson think that the dog bit?

These types of dependency create analytical challenges for an approach based on simple distribution and category: specifically, how to keep track of what the object of a verb is, or which DP is the subject of a passive verb. They require us to go beyond the Bloomfieldian systems of classification and distribution that are analysed as a simple phrase structure grammar, if we are to explain how different sentences relate to each other.

There are a number of approaches to tackling these examples of non-local dependencies.

For examples like (1) and (2), one can simply allow a phrase structure grammar to generate both active and passive verb phrases, capturing the generalization that these verb phrases, and the sentences that contain them, are related to each other by either redundancy rules in the lexicon (as in Lexical Functional Grammar, Bresnan 1982) or by rules that generalize over phrase structure rules (metarules, as developed in Generalised Phrase Structure Grammar, Gazdar et al. 1985). In either approach, the non-locality of the dependency is captured by relations stated outside the core structure building mechanisms of the grammar.

For examples like (3), which cross clauses, a lexical approach is more difficult to argue for. However, there are techniques that can be used in defining phrase structure rules which have the effect that information about dependencies can percolate through structure. That is, the selection and thematic role assigning properties of *bit* in (3) are transmitted through the structure until they connect with *which cat*, allowing an explanation of the fact that *which cat* is interpreted as the object of *bit*. The explanatory task, within such an approach, is to determine the principles of information transmission.

These approaches take a viewpoint where a syntactic representation is a single tree structure, annotated with information as to how the properties of the various elements in that structure connect. Phrase structure rules, appropriately annotated and augmented with a rich lexicon, can provide an account of how the non-locality of syntactic dependencies works in human language. However, the complexity of the linguistic techniques used to handle such phenomena is higher than one would like, raising issues of learnability.

An alternative is to develop a system where the noun phrase *Lilly* in (1) is actually present in two places in the structure: it is both the complement of *bite*, and the specifier of T. The same element structurally relates to those two positions. Similarly, *which* cat in (3) is the specifier of *did* and also the complement of *bit*. This means that a syntactic representation cannot be a simple phrase structure tree. This is because, in a phrase structure tree, one element cannot be immediately dominated by two distinct nodes, but if *Lilly* in (1) is the complement of *bite* and the specifier of T, it is immediately dominated by VP and TP. Such representations violate the No-Tangling Condition (Partee et al 1993, p440).

Early approaches to this second alternative therefore involve multiple levels of syntactic representation. The syntax of a sentence is not a single tree structure, but rather a collection of such structures, each related to the other by derivational steps: transformations. When a single element is in one syntactic position in one tree structure, and is in another in another tree structure, and these structures are related by a transformation, the phenomenon is typically referred to as Movement. For (1), *Lilly* is the complement of *bite* in one tree structure and this tree structure is related to another structure where *Lilly* is the specifier of T by a Movement transformation that connects the two positions in the two trees. Each structure is a well-formed phrase structure tree, but the representation of the sentence involves both.

In these early approaches, the movement transformations themselves played a major role in providing explanations of the phenomena of human language. However, the range of

possible transformations was vast, again raising learnability issues, so a theory restricting transformational types was needed, and was developed over the 1970s. By the time of the GB era, the transformational component of the grammar had been more or less reduced to a single rule (Move-alpha), which related tree structures. The explanatory work was done not by distinct transformations, creating particular constructions, but rather by more abstract conditions on what kinds of tree structure were legitimate at different stages of the syntactic derivation: X-bar theory held of all the structures; theta theory held of one, case theory of another, binding theory of yet another. The single movement rule effectively acted as a generator of relationships between tree structures whose well-formedness was determined by a series of constraints. In this theory, there are quite distinct theoretical devices responsible for generating structure (phrase structure rules constrained by X-bar theory), for constraining it (case theory, EPP, etc), and for relating structures to each other (Movement). Movement is a special device to capture non-local relationships.

A major departure in early minimalism was to remove this distinction: rather than having distinct devices to create structures and to relate them, Chomsky (1995) proposed a single set of rules that would both create and relate structures. This did away with the idea that there were sets of trees, and that syntactic representations were phrase structural. The initial proposal was that there were two core rules that built structure: one building structures locally (Merge) and another building non-local structures (Move). The structures built by Move violate the No Tangling Condition: a moved element is in two distinct places in structure, and so has two mother nodes. The structures built by Merge and Move, like phrase structures, are hierarchical, but unlike them, they are unordered and broadly set-theoretic in nature. Minimalism, then, rejects the idea that syntactic representations are collections of tree structures, an idea which had been a defining characteristic of transformational grammar in one way or another since before Aspects of the Theory of Syntax (Chomsky 1965). But rather than adopting the idea that these representations are single, richly annotated phrase structure trees, Minimalism took them to be dynamically constructed fractal-like objects, built through recursive application of the same simple computational processes.

However, the notion that there was a pair of such processes (Merge and Move) was also soon rejected, and Chomsky in 2001 (published as Chomsky 2004) proposed a further radical reduction in the range of theoretical devices, noting that a simple definition of Merge, allowing it to reuse structures that it had already created, provides an immediate account of the existence of movement in human language. This theoretical manoeuvre takes the definition of syntactic object to be as follows:

(4) a. A lexical item is a syntactic object
b. If A is a syntactic object, and B is a syntactic object, then Merge(A, B) = {A, B} is a syntactic object.

Chomsky notes that there are three logical subcases here. B is distinct from A, B is part of A, or B is A. The first subcase, which he terms External Merge, would correspond to non-movement generated structures (analogous to those created by phrase structure rules); the second, which he terms Internal Merge, corresponds to movement structures, since, if B is part of A, it is contained within A, and if it is Merged with A, it is not contained within A. This

means that B is in two distinct places in a single structure. The final case involves A and B being identical (so called Self-Merge), and may be restricted to the initial stage of derivations (e.g. Kayne 2008) or be used as a means of creating unary projections (Adger 2013).

The important conceptual point here is that, in current Minimalist syntactic theory, there is no longer any distinction in the mechanisms used to build structures involving local vs non-local dependencies. The previous distinction, captured in the different mechanisms of phrase structure and transformations (or, similarly, X-bar Theory and Movement), has been erased, and a single, unified mechanism is responsible for creating all syntactic structures. Unlike the approaches to non-local dependencies in phrase structure grammar, which essentially eliminate transformational relations in favour of phrase structural ones, Minimalism effectively eliminates phrase structural relations in favour of transformational ones, using the core mechanism for movement as a structure building operation.

References

Adger, D. (2013) *A Syntax of Substance*. Cambridge: MIT Press.

Bloomfield, L. (1926) A set of postulates for the science of language. *Language*, *2*: 153-164.

Bresnan, J. (1982) The Passive in Lexical Theory. In J. Bresnan, ed., *The Mental Representation of Grammatical Relation,* Cambridge: MIT Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti, ed., *Structures and Beyond*. 104-131. Oxford: Oxford University Press.

Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized phrase structure grammar*. Cambridge: Harvard University Press.

Kayne, Richard S. 2008b. Antisymmetry and the lexicon. *Linguistic Variation Yearbook 8*: 1-31.

Partee, B. B., ter Meulen, A. G., & Wall, R. (2012). *Mathematical methods in linguistics.* Dordrecht: Kluwer Academic Publishers.