Chapter 1

Merge and Features: The Engine of Syntax[1]

Peter Svenonius

CASTL, Department of Language and Culture, University of Tromsø – The Arctic University of Norway, N-9037, Tromsø, Norway

*peter.svenonius@uit.no

Abstract: Merge is the central engine of syntax, in Chomsky's Minimalist Program. In this chapter I discuss Merge, which is relatively well understood, and formal syntactic features, which are much less well understood, and their relation to each other in Chomsky's work and related efforts.

Keywords: Merge, Agree, Features

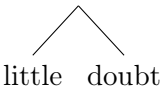
1.1. Introduction

Since the inception of the Minimalist Program (Chomsky 1993, Chomsky 1995b), Chomsky has increasingly pushed the boundaries of a minimalist conception of grammar, one which pares the mechanisms specific to language down to the bare bones. Central to this minimalist conception is the notion of Merge. Merge creates complex syntactic objects from simple lexical items: *little* and

[1]Thanks to Nicholas Allott, Terje Lohndal, Kristine Bentzen, and Noam Chomsky for comments on earlier drafts.

doubt together make *little doubt*.

(1)

LEXICON	+ MERGE	= STRUCTURE
little	⇒	
doubt		
...		

In addition to the structure-building operation of Merge, grammar involves some mechanism for feature interaction. Natural language agreement and concord phenomena and case patterns demonstrate this clearly, for example the sentences in (2a) and (2b), where *she* and *her* are different forms of the same pronoun, conditioned by syntactic structure, and similarly for *they* and *them* and *seem* and *seems*. Note also in (2c–2d) that agreement is possible without raising.

(2) a. She seems to them to be absent

b. They seem to her to be absent

c. There seems to them to be somebody absent

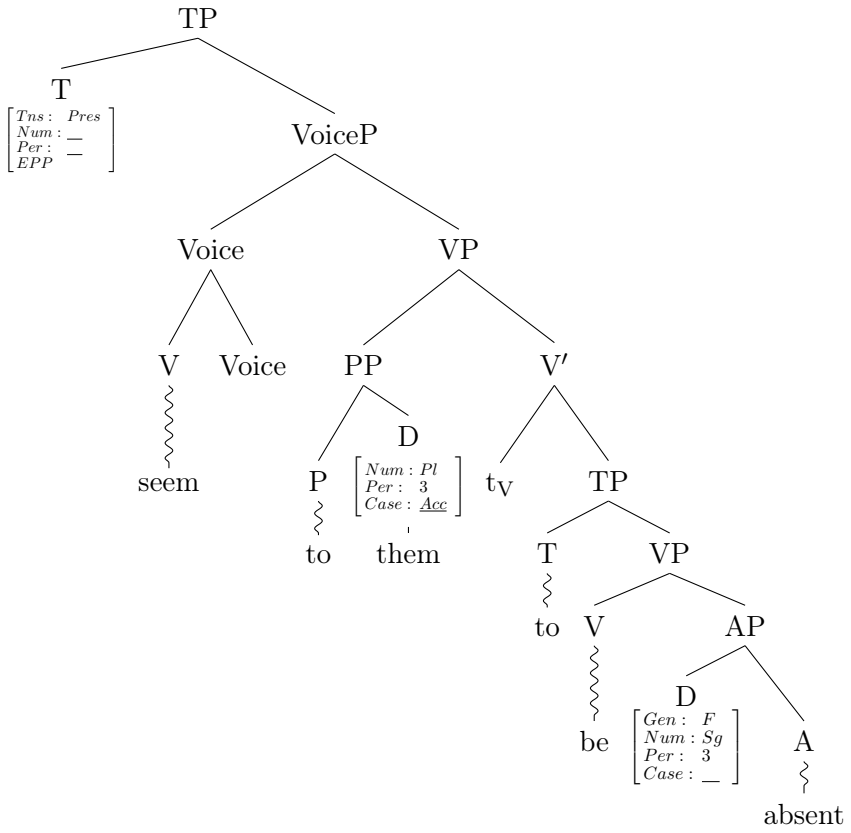
d. There seem to them to be people absent

In these sentences, the subject gets its thematic interpretation from the lower, infinitive clause, before raising in (2a–2b) to become the subject of the finite clause. In the tree in (3), corresponding approximately to the pre-raising stage of (2a), the pronouns are represented by Ds with different features, and the inflection on the finite verb is represented by the feature bundle in the matrix clause (other features and nodes are omitted for purposes of presentation).[2] I depict the main verb *seem* as having undergone 'head movement' to a Voice head to precede the prepositional argument *to them*. Two nodes have incomplete

[2]Standard abbreviations for syntactic categories used here include D[eterminer] (including pronouns), V[erb], T[ense] (for clauses), A[djective], and in later trees also C[omplementizer] (for clauses extended by a complementizer, a subordinator or a marker of illocutionary force).

feature specifications at this stage of the derivation and so I have not associated them with phonological exponents[3] — the subject pronoun and the finite tense.

(3)



Chomsky (2001, 4) notes several things that must be accomplished by a theory of feature interaction. For one thing, the inflectional head T of the matrix clause must copy ϕ-features[4] in order to express agreement. Here, this is accomplished through unspecified values for Num[ber] and Per[son] on T, as indicated with blanks. For another thing, that agreeing relation must skip the argument in the prepositional phrase to find the more distant one in the embedded infinitival complement. (Chomsky, 2000, 123) suggests that the presence of unvalued features makes a node visible ('active') for feature interactions — and

[3] An exponent is the phonological information associated with a syntactic head, here connected to the syntax with squiggly lines.

[4] Chomsky (1981, 330) identified person, number and gender as the formal features which constitute pronouns and traces of movement, and dubbed them 'ϕ-features'.

the D in the PP has its case valued by P, hence has no unvalued features, while the subject of the infinitival does not have its case valued until it enters into an agreement relationship with the matrix T. Third, finite T requires a subject; in the case of (2a), the D with which it agrees must raise) (in (2c–2d), the indefinite noun phrases can remain low, and an expletive subject is inserted). This is accomplished through a feature called "EPP" (for historical reasons; see Chomsky 1982, 10).

Chomsky has explored several different varieties of a mechanism AGREE (Chomsky 2000, 101) responsible for these kinds of feature interactions, for example feature matching, feature valuation, and feature movement, as discussed further below.

I discuss Merge in §1.2, Agree in §1.3, and the possibility of subsuming Agree to Merge in §1.4.

1.2. Merge

1.2.1. The necessity of Merge

In order for a language to be creative, in the sense of being able to encode content which has not previously been encountered, it must have a grammar, a system for combining lexical items (Chomsky, 1964, 1966). Grammars can be stated in terms of constraints on representations (see Sells this volume), but sentences are not listed—any speaker of any natural language can spontaneously create utterances which have never before appeared. A derivational model directly manifests the fact of creativity. Architectural assumptions such as the choice between derivation and representation may also have important consequences for the form that explanation takes (see Adger this volume for an example).

In a derivational model, a grammar must include at least a lexicon and a set of operations. Transformational Grammar had phrase structure rules to generate structure, first piecemeal (Chomsky 1965) and then more generally under the $\overline{\text{X}}$ (X-bar) schema (Chomsky 1970). In the Minimalist Program, the basic operation combining lexical items is called MERGE (Chomsky 1995a, 62,

Chomsky 1995b, 226).

(4) Applied to two objects α and β, Merge forms the new object γ. (Chomsky, 1995b, 396)

Merge, in addition to applying to lexical items, can also operate on linguistic objects which have been created by previous applications of Merge, a property which introduces the recursive nature of language.

The Minimalist Program leads to the question of whether Merge could be sufficient for Narrow Syntax (the generation of structures which can be interpreted at the Conceptual-Intentional interface); that is, could the sole operation of Narrow Syntax be Merge, with all other aspects of syntactic computation being derived from interface conditions and properties of lexical items?[5]

One can imagine a very simple language consisting entirely of compounds, with free adjunction of lexical words as in *banana-give*. As a thought experiment we can try to imagine what properties such a language would develop in interfacing with systems of use. In practice, actual implementations of even simple grammatical descriptions of real natural languages always appear to require more than Merge, but in this chapter I will explore some of the ways in which various different mechanisms might be reduced to Merge, in the spirit of the Minimalist Program.

[5]This has been stated in various forms as a working hypothesis, for example in Hauser et al. (2002, 1573): "FLN [the faculty of language in the narrow sense, "the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces"] comprises only the core computational mechanisms of recursion as they appear in narrow syntax and the mappings to the interfaces". Chomsky (2008, 139) refers to Merge, again in the context of a hypothesis, as "the sole computational operation of narrow syntax". Chomsky (2010, 52) characterizes the *Strong Minimalist Thesis* as 'Interfaces + Merge = Language', a thesis to be tested and evaluated. In the same vein, Chomsky et al. (2019, 237) write "given that [Merge] is what is minimally necessary to create hierarchical structure, we assume that this is the *only* operation defined by UG", though they mitigate this in a parenthetical alluding to Pair Merge, on which see below. See Fitch et al. (2005) for pushback against the characterization of their (Hauser, Chomsky, and Fitch's 2002) position as claiming that Merge is all there is to FLB, the faculty of language in a broad sense. See also Adger this volume on the difference between the syntactic computation and, for example, processing.

Chomsky describes Merge in terms of set formation, on grounds of simplicity (e.g., Chomsky 2005, 15–16). On that conception, two elements which are merged form a set, which does not in itself establish any asymmetry between the two (e.g., {*little, doubt*}).

(5) γ is the set $\{\alpha,\beta\}$ (Chomsky, 1995b, 396)

Sets are mathematically fundamental (Russell and Whitehead, 1913). This does not necessarily mean that set formation is the best model for the ultimately biological processes underlying Merge; computational systems are not set based.[6] However, in the absence of clear evidence to the contrary, the simplest hypothesis appears to be that Merge creates a set.

In $\overline{\text{X}}$ theory, endocentricity (headedness) was built in in the form of 'Projection' of a category from the head to the phrase level. This was carried over into early formulations of Merge as follows (where α, β, and γ can be read as referring back to (4–5)).

(6) "...verbal and nominal elements are interpreted differently at LF and behave differently in phonological component ... γ must therefore at least (and, we assume, at most) be of the form $\{\delta, \{\alpha, \beta\}\}$, where δ identifies the relevant properties of γ; call δ the label of γ." Chomsky (1995a, 62)

(7) "The operation Merge, then, is asymmetric, projecting one of the objects to which it applies, its head becoming the label of the complex formed." Op cit., 63

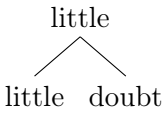
(8) "one or the other of α and β projects and is the head of γ" (Chomsky, 1995b, 397)

[6]For a freewheeling discussion of Merge and set theory, the interested reader may wish to consult a post by Norbert Hornstein on the *Faculty of Language* blog on February 12, 2017 and the ensuing discussion in the comments section of that blog. For example comments by Thomas Graf on February 26th regarding the inappropriateness of sets for computational applications are relevant here.

Chomsky models this in set-theoretic terms, with one of the two elements forming the label for the whole, as depicted in (9) on the reading where *little* is a quantifier specifying a quantity of doubt (as in, *I have little doubt*).

(9) {little, {little, doubt}}

Represented in tree form, a structure is labeled with the label of its head.

(10)


Chomsky (1995b, 248) also proposes another operation, Pair Merge, which creates asymmetric adjunction structures. This might be involved in the other reading of *little doubt*, where *little* is an adjective restricting the size of a doubt (as in, *a little doubt grew into a big one*).

See Lasnik and Uriagereka (2012) for discussion of the adjunction structures that motivated Chomsky to propose Pair Merge as an operation distinct from Merge, which is sometimes called Set Merge to distinguish it from Pair Merge. Hornstein (2009, ch. 4) proposes instead that Merge be decomposed into concatenation and labeling (see §1.3.3 below on labels), and that adjunction structures are unlabeled. In that case something like Pair Merge is a subpart of Set Merge.

1.2.2. Movement as Internal Merge

In addition to the Merge of distinct syntactic objects, natural language also exhibits movement, in the sense that there are syntactic objects which must be interpreted in more than one position, e.g., a thematic position like the underlined positions in (11) and a discourse-informational or operator position like the clause-initial positions in (11b–e).[7]

[7]The fundamental uniformity of these different constructions, developed especially in Chomsky (1977), is a major discovery of generative linguistics.

(11) a. They told us to put the pills in cups.

b. Which pill did they tell us to put __ in which cup?

c. The red pill, they told us to put __ in the tall cup (and the blue pill in the short cup).

d. the pills that they told us to put __ in the cups

e. as many pills as they told us to put __ in the cups

In early transformational grammar (e.g., Chomsky 1965), the base component building phrase structure was thought to be distinct from the transformational component which included movement (similarly, representational theories of grammar use specialized devices like SLASH or conditions on chains to characterize filler-gap dependencies like those in (11b–d)). but in the Minimalist Program, the existence of movement, and its properties, fall out as a consequence of the way Merge works. Chomsky (2004) proposed to unify movement with Merge by allowing Merge to relate a syntactic object to a subpart properly contained inside that same syntactic object; this is called INTERNAL MERGE (Chomsky 2004, 110; see also Bobaljik 1995, Starke 2001).

This substantially increases the importance of Merge as a central component of grammar, and simplifies the inventory of mechanisms, very much in keeping with the aims of the Minimalist Program.

There are at least two ways to think about movement which are consistent with Internal Merge. One is in terms of copies (Chomsky 1993; Nunes 1995): the internally merged element exists in two copies which form a "chain". The higher (moved) copy has all the properties of the lower (original) one except its position in the tree, i.e., its set of dependencies.

(12) a. x . . . x formed by application of internal Merge is a chain; only one occurrence of x is spelled out (the higher, for overt movement)

b. *Who* did they ask *t*?

c. D (spelled out as *who*) . . . D (its trace) is a chain

A major accomplishment of generative linguistics is the discovery that movement is punctuated by clause boundaries, or 'successive-cyclic' Chomsky

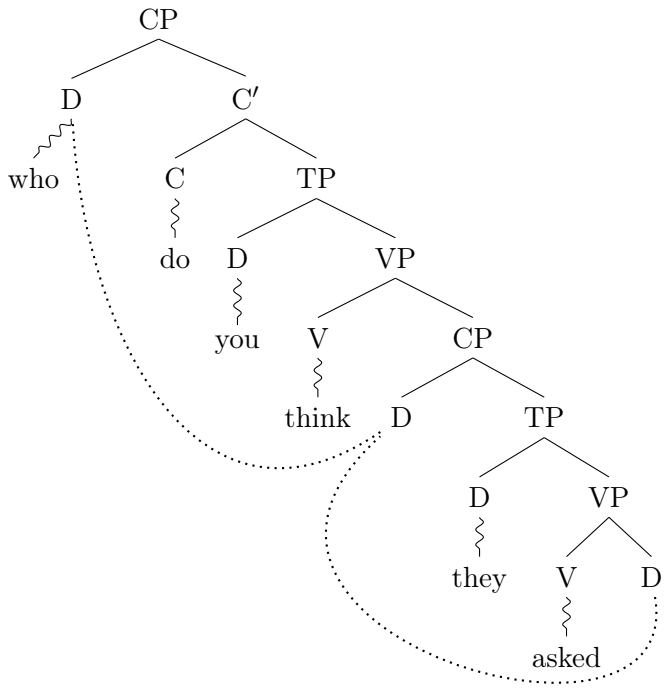
(1973), strongly supported in subsequent work such as McCloskey (2002). McCloskey shows that in Irish, the form of the complementizer directly reflects whether a moved element has passed through the edge of an intermediate clause, providing direct evidence that movement which appears to be'unbounded' in the sense that it may cross multiple clauses actually occurs in shorter steps of approximately clausal size.

In terms of copy theory, a copy is available for a new round of Merge, so that a chain may have several links.

(13) *Who* do you think t they asked t?

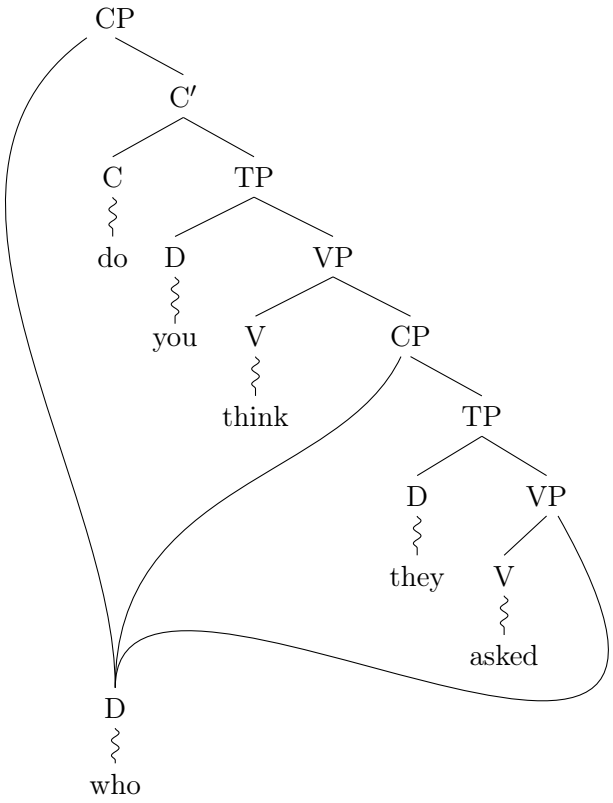
In the tree in (14), the dotted line is the chain. The tree has been simplified to show only heads which are overtly spelled out. I am setting aside the additional landing site at the edge of the verb phrase proposed by Chomsky (1986a) and Chomsky (2001).

(14)



A second way to think about movement is in terms of multidominance (Starke 2001; Gärtner 2002). A constituent of a syntactic tree is remerged, or internally merged, with the root of the tree containing it. Thus a single object enters into two sets of dominance relations, which determine what nodes it c-commands (a node c-commands what it dominates, and its sister, and all nodes dominated by its sister). There is a set of nodes which are c-commanded by an occurrence of an element in a low position, and an expanded set is c-commanded by the occurrence in a higher position. Multidominance is depicted in (15) (as above, simplifying the tree to only show nodes dominating overt material).

(15)



In (15), there is no copy, hence no chain. Instead of copies, three dominance relations for the D *who* are simultaneously represented. Chomsky (2019) observes that such structures violate a 'Single Mother Condition'. He suggests

that copies fall out naturally from minimal assumptions about Internal Merge, and that multidominance requires complications.

1.2.3. The central role of Merge

Movement plays an outsized role in generative grammar, including not only the derivation of so-called "unbounded" dependencies (overt A-bar movement, as illustrated for example in (12b) and (13) in §1.2.2, unbounded in the sense that they can pass through any number of finite clauses), but also A[rgument]-movement, head movement, covert movements of various types, and possibly others. The extension of Merge to encompass movement establishes Merge as the central operation in grammar, responsible for structure building.

Structure is fundamental to compositionality, which is the backbone of semantic interpretation. Wholly determined by the order of Merge are the differences between (16a) and (16b) and between (17a) and (17b).

(16) a. Jones promoted the conspiracy.

b. The conspiracy promoted Jones.

(17) a. The counsel didn't establish that the defendant did lie.

b. The counsel did establish that the defendant didn't lie.

Structure also determines domains for binding, for example where c-command constrains coreference. In example (18a), *rags* is embedded in the complex subject *linseed oil in rags* and as a consequence does not c-command *them*, with the result that the two may be coreferent. In (18b), in contrast, coreference with *rags* is only possible with a reflexive, because a non-reflexive pronoun cannot be coreferent with an expression that c-commands it in the same clause (Chomsky, 1980).[8]

[8] I depict *rags* in (18b) as the locus of reference to highlight the asymmetry with (18a), but the same facts hold if it is the complex expression *rags in linseed oil*, with *rags* as its head, which bears reference.

(i) [Rags in linseed oil]$_i$ can ignite {themselves$_i$/*them$_i$}.

(18) a. Linseed oil in $[\text{rags}]_i$ can ignite $\{\text{them}_i/*\text{themselves}_i\}$.

b. $[\text{Rags}]_i$ in linseed oil can ignite $\{\text{themselves}_i/*\text{them}_i\}$.

Internal Merge plays a critical role in configuring material for semantic interpretation in the theory of covert movement (Huang (1982); May (1985)). Covert movement is exactly like ordinary (overt) movement except that instead of the higher copy being pronounced, the moved element is pronounced in its lower position, as if it had not moved. This means that the position for compositional interpretation is partly dissociated from the position of 'surface' pronunciation.

An example is a question with a wh-expression *in situ*, i.e., apparently unmoved, such as the French *où* 'where' in (19).

(19) *Tu vas où?*
you go.PRES.2SG where

'Where are you going?'

If *où* moves covertly, then French-like languages can be unified with English-like languages. Potentially, all wh-questions require a wh-operator to move to a higher layer in the clause, allowing the same rules of semantic interpretation to apply to (19) (and to the inverted *Où vas-tu?*) and to English *Where did you go?*. This points toward a universal compositional mapping from syntactic structure to semantic interpretation. Where languages vary is in the mapping between syntactic structure and phonological expression. Chomsky argues that language privileges thought over communication, systematically simplifying the semantic computation at the expense of pronunciation or 'externalization' (Chomsky (2005, 4), Chomsky (2007, 12–15), Chomsky (2010, 15–16), Chomsky (2013a, 654–660), Chomsky 2017).

Merge also plays a key role in word formation, in some implementations of the theory. Yang et al. (2017) mention cases of structural ambiguity at the word level, which can be exemplified here with *[over-fragment]-able* ('susceptible to being too much fragmented') versus *over-[fragment-able]* ('too much susceptible to being fragmented'). Internal Merge plays an important role as well, in the theory of head movement (Baker (1988), inter alios). There, a complex word

like the French *va-s* 'go.PRES-2SG' in (19) is argued to be formed by syntactic movement (i.e., Internal Merge) of the verb to an inflectional head (e.g., Embick 2010, Harley 2014).

Somewhat more radically, it has even been suggested that Merge is responsible for combining phonological segments into syllables and syllables into larger prosodic structures (Yang et al. 2017).

These various extensions of Merge (covert movement, word formation, and phonological structure) are controversial, but they illustrate different aspects of the potential power of Merge as an analytic tool. No rival unification of semantics, morphology, and phonology is visible on the horizon.

As a final example of the ways in which Merge has been extended to cover phenomena beyond the construction and displacement of phrases, I discuss in §1.4 below the hypothesis that feature checking itself is a kind of movement (feature movement, Chomsky 1995b, 262–271, Taraldsen 1996, Ochi 1999, Pesetsky 2000).

1.2.4. Free Merge and Triggered Merge

A question which has been debated recently is the question of to what extent the operation Merge interacts with features of the Narrow Syntax, perhaps being 'triggered' by a formal requirement (Pesetsky and Torrego 2006), or alternatively 'free' and filtered only at the interfaces with other systems (Boeckx 2010). The matter is complex and unsettled.

Traditionally, lexical heads are thought to have argument structure as a syntactic property, for example verbs can be transitive or intransitive, which can be modeled if the transitive verb has a 'selectional' feature which requires it to merge with an object, and the intransitive verb lacks such a feature, and Merge is not licensed in the absence of selectional features (Emonds 2000).

(20) a. The scientists defeated the virus.

b. * The scientists defeated. (*defeat* is transitive, requires DP)

c. The scientists prevailed.

d. *The scientists prevailed the virus. (*prevail* is intransitive, can't com-

bine with DP)

Functional complementation, for example when the functional T takes a verb phrase complement, or when the functional D takes a noun phrase complement, has been taken to work the same way (Abney 1987).

(21) a. the dog

b. *the (*the* is "transitive" (requires NP))

c. him

d. *him dog[9] (*him* is "intransitive" (can't combine with NP))

However, the argument structure of verbs is flexible and subject to context and other factors, leading some to argue that argument structure is not regulated by selectional features but by conceptual factors (Borer 2005), which are not categorical. This would mean that there is something about the concept encoded by *defeat* that makes it ill-suited to intransitive use, while *prevail* is conceptually intransitive, but that Narrow Syntax has nothing to say about that.

This has led to the possibility that Merge in general is not triggered by syntactic features (Boeckx 2014; Chomsky et al. 2019). It is conceivable that something about the conceptual content of *the* requires it to appear with a noun phrase, and something about the conceptual content of *him* prevents it from doing so; and that this is adjudicated at the interface, beyond the purview of Narrow Syntax. However, this pushes much of what is usually thought of as grammar out of the Narrow Syntax and into the interfaces, highlighting the need for a theory of grammar which includes more than Narrow Syntax.

It is possible to accept Borer's arguments that the contrasts in (20) are not syntactic but to continue to maintain that the contrasts in (21) are.[10]

[9]This is bad as a determiner phrase, not to be confused with a compound like *he-dog*, which would have stress on *he* and could refer to a male dog or a masculine dog. Compare the phrasal *you dogs*, where it could be argued that the pronoun *you*, unlike *he* and *him*, is optionally transitive.

[10]In that case (20b) and (20d) should actually be marked with crosshatches to indicate semantic anomaly, rather than asterisks to indicate ungrammaticality.

1.3. Features

1.3.1. Features in syntax

Chomsky (1965, ch. 2) (*Aspects*) introduced features into generative syntax to capture generalizations invoked by grammatical rules that cross-cut syntactic categories, and the system of features was subsequently developed in important ways (see Muysken and van Riemsdijk 1986 for the state of the art up to that time).

For example, *where* and *what* both have special properties as wh-expressions, fronted in interrogatives, but *where* corresponds to a prepositional phrase gap (*Where do they live? They live in tents*) while *what* corresponds to a noun phrase gap (*What do they make? They make tents*).

A major issue in *Aspects* which has persisted was to distinguish feature-based constraints holding in the syntax from conditions based on conceptual semantics, which might not be feature based and might not belong in the syntax at all.

1.3.2. Agree

Since the appearance of Chomsky (2000) and Chomsky (2001) there has been an increased interest in developing a minimalist theory of feature interactions (see Adger and Svenonius 2011 for an overview and references).

As already mentioned in §1.1, natural language agreement for ϕ-features provided the model for the operation AGREE (Chomsky, 2000). The basic idea behind Agree is that two nodes in a tree can share values for a feature F if one is a PROBE for F and the other bears F and is sufficiently local to the probe — these facts make it a GOAL.

Thus the two nodes entering into the Agree relation share something from the outset, namely F, which identifies what kind of information they can share, but are also distinguished in that one of them is a probe, causing it to seek a goal. Adger and Svenonius (2011) call the property that distinguishes a probe from a non-probe a 'second order property' of a feature.

This basic model of feature checking has been developed in many guises. It

has been extended to handle selection, movement, licensing, and other operations.

For example, if the verb *live* (in the sense 'reside') selects a PP complement, this could mean that it has a probe for a feature borne by PPs (see Svenonius 1994 for an implementation). If a negative polarity item like *ever* must be licensed by negation, it could have a probe for a negative feature (Zeijlstra 2008). A case-marked noun phrase requires a case assigner, so a noun phrase could probe for a class of features carried by case-assigners (see Pesetsky and Torrego 2001 on modeling case assignment with Agree). Probes can be specified as to whether they require movement of their goal or not (the second order property of 'strength', cf. Chomsky 1995b, 232), and whether failing to find a suitable goal leads to ungrammaticality or just the realization of a default value.

Feature checking theory promises to unify agreement, selection, movement, and licensing all in a fairly simple mechanism. This is formally implemented in minimalist grammars like that of Stabler (1997).

In Chomsky (1995b, 280ff) and Chomsky (2000), the second-order property of probes which caused them to trigger operations was the property of UNINTERPRETABILITY. The core idea was that a semantically interpretable feature like number, for example expressing the plurality of *dogs*, could have an uninterpretable counterpart, for example in a verb form like *are*, which has the same meaning as singular *is* (for example, a plural verb form does not imply anything about the number of events). The interpretation comes from the interpretable instance of the feature on the noun which controls the agreement.

Chomsky (2001, 5ff) proposed that probes were necessarily unvalued, seeking to copy a value from a more fully specified goal, as illustrated in (3). Pesetsky and Torrego (2007) argued that interpretability and valuation were dissociated, and that both were independently needed.

1.3.3. Labels

Chomsky (2008, 2013b, 2015) argues that the imposition of endocentricity on Merge (see (6–8) in §1.2) is unneeded. Instead, Merge is taken to be fundamentally symmetric, creating unlabeled, unheaded sets. The source of headedness

in syntactic phrases is not Merge, but is to be found at the interface. Specifically, Chomsky, suggests, syntactic objects must be labeled in order to be interpreted at the Conceptual-Intentional interface. Accordingly, an operation LABEL is introduced which creates a label for a syntactic object before it is interpreted (see also Hornstein 2009 and Adger 2013).

The theory of labeling which is advanced in these papers, especially Chomsky (2013b), has important consequences for the theory of features.

In the simplest case, one of the two elements merged is a head, and its label is copied by Label, through "minimal search". Hence, a V merged with an argument is labeled V (VP by convention), a T merged with a predicate is labeled T(P), and so on, for the most part giving the same result as the old theory of Merge-with-projection — or, indeed, $\overline{\text{X}}$ theory, Chomsky (1970).

But Label is not some kind of notational variant of $\overline{\text{X}}$ theory. As Chomsky (2013b, 46) notes, the Label theory has the consequence that pronouns like the ones in (3) cannot be simple heads, as they are depicted there, but must be syntactically complex. If they were simple heads, they would provide labels for the phrases that they merge with, so that the PP *to them* and the AP *[her] absent* (or the TP *[she] seems to them to be absent*, after raising) would then have the distribution of noun phrases, contrary to fact.

When two phrases are merged, minimal search does not produce an unambiguous label. A central intuition of checking theory was that the probe and the goal share a feature (as noted in §1.3.2). In the same spirit, Chomsky suggests that Label can use a shared feature as the label for the whole. Thus, if a tense head bears ϕ features and it is merged with a subject DP bearing matching ϕ features, the resulting clause can be labeled with those ϕ features (recalling the era of Chomsky (1986b), when the clause was a projection of INFL for 'inflection'). Similarly, if an interrogative CP bears a Q feature and so does a wh-pronoun which merges with it, then the resulting structure can be labeled with the Q feature (see Rizzi (2015) and Chomsky (2015) on deriving 'freezing' effects from this).

On this view, no second order probe property need necessarily be involved. If the interface demands labels, and Merge is free, then a structure in which a DP has been merged with a predicate will be unlabeled and hence fail to be interpreted at the interface.

Consider the example of a small clause (like the AP in (3)). Moro (2000) argued that simple subject-predicate combinations were syntactically unstable because they could not be linearized at the phonological interface, forcing the subject or the predicate to raise. Chomsky (2013b) adopts the notion that small clauses are unstable and require movement, but identifies the instability as being a lack of a label, locating the problem at the Conceptual-Intentional interface.

The DP can move up, but will create the same problem at a higher position unless it finds a node bearing ϕ features with which to merge. The subject may raise to the edge of a phrase with which it shares features, such as INFL or T.

These considerations are sufficient to force the raising of the DP subjects in (2a–2b) in §1.1 to the finite TP, since the embedded nonfinite T has no ϕ features. They generalize to all clauses with an agent or actor subject, because those have been persuasively argued to be merged with a verb phrase (not simply a verbal head). But there are other sentences in which an adequate low position exists for a DP, and nevertheless something must be inserted in subject position, as in (2c) and (2d) — in English, though not in Italian.

Chomsky (2015) suggests that in null subject languages like Italian, the rich inflection in T provides a label for for the object it forms by merging with a verb phrase, but in English, with weaker inflection, T is defective, and cannot label TP unless it is 'reinforced' by an overt subject. Thus the EPP feature in (3) is replaced by a theory of weak label features. In the terms of Adger and Svenonius (2011), weakness would be a second order property of the features of English T.

Chomsky also discusses another important case in which phrasal Merge does not involve shared features. That is the successive-cyclic movement of an operator to an intermediate landing site (as with the first step of movement in (14)).

An example like (13), namely *Who do you think they asked?* starts out with an embedded clause *C they asked who* (assuming all embedded clauses are introduced by C). The C is not interrogative and the wh-expression cannot be interpreted in this clause. Movement (Internal Merge) is free, so the wh-expression can move to the edge of the clause, but it shares no features with C, so the result cannot be labeled. However, a subsequent step of movement to

the matrix C, which is interrogative, results in a labeled structure, and once the offending wh-expression has evacuated the embedded clause, the embedded C is free to label the phrase containing it (essentially, traces of movement are ignored by the Label operation).

The fact that movement can never skip an intermediate clause boundary altogether is due to the fact that each clause is separately packaged for interpretation at the interfaces (Chomsky, 2001); this much is constant for the treatments being discussed here.[11] What has shifted is the status of non-feature-driven movement.

The theory of Chomsky (2000, 2001) involved rich featural interactions, as represented in (3). At that stage in the theory, it seemed plausible to imagine that all operations were driven by features, and by second order properties of features like the probe property. This makes Agree a central component of narrow syntax, alongside Merge (as depicted in the diagram of Minimalist grammar architecture in Alexiadou and Lohndal's (18), this volume).

However, no independent motivation was found for a feature that could drive the non-final steps of successive-cyclic movement. Subsequent developments, in particular Chomsky (2008, 2013b), change that picture substantially in dialing back the expectation that movement should be driven by second order properties of features — in other words, in the terms of §1.2.4 we move from the expectation that Merge be triggered to the expectation that it be free.

Free merge changes the status of intermediate stages of movement. Movement to final landing sites is related to features, but not by second order properties like strength, or diacritics like EPP. Rather, the interface requires a label, and the algorithm providing labels requires a feature match between the moved element and its host. This has the potential to greatly simplify the theory of

[11] The gist of Chomsky's (2001) PHASE theory is that when C is merged, its complement TP is interpreted. Assuming that a wh-expression must be associated with an appropriate C at the time of interpretation, then a TP which contains an unassociated wh-expression will not be interpretable, unless the wh-expression moves out of it, to the nearest available C — the one which triggered the spell-out of TP (and possibly not the final C with which the wh-expression will be associated). At that stage of the derivation, there are no higher positions.

features, reducing the explanatory burden on valuation, interpretability and other aspects of the theory of Agree. In fact, Chomsky et al. (2019, 252) hint at the possibility of eliminating Agree from the Narrow Syntax, relegating it entirely to the interface.

A challenge, however, is provided by languages like Irish as described by McCloskey (2002). Recall from §1.2.2 that Irish complementizers morphologically express a featural distinction between C which is targeted by intermediate steps of successive-cyclic movement and C which is not. This provided one of the more striking confirmations of the successive-cyclic nature of movement. But it also seems to be at odds with the feature-free account of intermediate steps of movement.

1.4. Feature movement

The previous sections have been concerned with issues that have been discussed hotly in recent Minimalist approaches to language. In this section I turn to an issue which does not seem to currently be at the forefront of the research agenda but which I think could have important impact.

Specifically, I briefly discuss the possibility, raised in Chomsky (1995b, 262–271), that the residue of Agree could be unified with Merge. For example, rather than T copying ϕ from D in a structure like that in (3), ϕ could move from D to T (either as a copy, or under multidominance, as discussed in §1.2.2).

Movement can be thought of as the creation of a dependency between the head at the root (e.g., interrogative C) and a projection inside the tree (e.g., a D with a [wh] feature). Feature-checking is also the creation of a dependency from one part of the tree to another part, e.g., unvalued ϕ on T and valued ϕ on D (see Frampton and Gutmann (2000) for a formalization of this perspective).

The idea was explored somewhat in the wake of Chomsky (1995b), along with other ways of unifying Merge and Agree (as in Hornstein 2009), but has not established a broad consensus. Occasionally feature movement has been explored as an alternative to covert movement, but Pesetsky (2000) argues that the two must be distinguished.

The mechanism of Agree was modeled on agreement. One problem with

identifying agreement too closely with movement is that ϕ-agreement does not show the properties of the minimal ϕ-bearing elements that actually move, namely pronouns (see Bresnan and Mchombo 1987 for an early and classic demonstration, and Rezac (2010) and references there for more recent examples). However, Cardinaletti and Starke (1999) and Déchaine and Wiltschko (2002) have argued that there is a typology of pronouns with different levels of referentiality. Conceivably, ϕ-feature bundles involved in Agree could be lower on a referential hierarchy than other pronominal elements, leading to their failing the tests suggested by Bresnan and Mchombo and others. In that case, agreement might be the movement of a particularly small subpart of D.

Another difference between agreement and movement is that agreement commonly results in the same features being spelled out in two places, while movement ordinarily does not. However, given that Narrow Syntax is the derivation of structures which are legible to the Conceptual-Intentional interface (cf. §1.2.1), the spell-out of movement chains (or of multidominance structures) is a matter of externalization and may not reflect a very deep difference between the two phenomena.

The rise of non-feature driven movement in the theory of Chomsky (2013b) and associated papers has reduced the importance of Agree, but has not eliminated it entirely. If the residue of Agree could be fully unified with Merge, that would reduce the number of operations in play in Narrow Syntax, fully in keeping with the overall thrust of the Minimalist program.

1.5. Conclusion

Ever since he started developing linguistic theory in the 1950's, Chomsky has displayed an impressive ability to revise his positions and break with the consensus up to that point, even when the consensus is built on a previous version of his own theory. With each turn, some part of the linguistic community follows Chomsky on the new path, and another part sticks by his earlier position. This was noted by Sells (1985) but has also held true since then.

The issues discussed in this chapter are no exception to that general pattern; Chomsky has held different positions at different times on the nature of Merge,

the nature of Agree, and all of the other matters raised here, and for each position there is a camp of adherents who don't think the subsequent moves were on the right track.

Adequate grammatical description of a natural language in the form of a written grammar requires thousands of pages of detail. But much of this concerns the properties of lexical and functional items, for example the properties of *which* or *since* or *as*. One consistent point throughout Chomsky's Minimalist Program is the pursuit of the hypothesis that the grammatical component is very simple, and it is only through the interaction of this simple grammar with the individual properties of lexical items that a complex grammar emerges.

In this chapter I have explored some ways in which the ideal of a maximally simple grammatical component has been approached. The most indispensible grammatical operation is Merge (§1.2). The features of lexical items which differentiate them from each other syntactically interact with grammatical operations, which may include Agree (§1.3). However, Agree might be partially or even wholly subsumed under a suitably generalized notion of Merge (§1.4).

Bibliography

Abney, Steven. 1987. *The English Noun Phrase in its Sentential Aspect*. Ph.D. thesis, MIT, Cambridge, Ma.

Adger, David. 2013. *A Syntax of Substance*. MIT Press, Cambridge, Ma.

Adger, David. to appear. The architecture of the computation. This volume.

Adger, David and Peter Svenonius. 2011. Features in minimalist syntax. In *The Oxford Handbook of Linguistic Minimalism*, edited by Cedric Boeckx, pp. 27–51. Oxford University Press, Oxford.

Alexiadou, Artemis and Terje Lohndal. to appear. The current state of generativism: From Government and Binding to Minimalism. This volume.

Baker, Mark C. 1988. *Incorporation: A Theory of Grammatical Function Changing*. University of Chicago Press, Chicago.

Bobaljik, Jonathan David. 1995. In terms of Merge: Copy and head movement. In *MIT Working Papers in Linguistics 27: Papers on Minimalist Syntax*, pp. 41–64. MITWPL, Cambridge, Ma.

Boeckx, Cedric. 2010. A tale of two minimalisms: Reflections on the plausibility of crash-proof syntax and its free merge alternative. In *Exploring Crash-Proof Grammars*, edited by Michael T Putnam, pp. 105–123. John Benjamins, Amsterdam.

Boeckx, Cedric. 2014. *Elementary Syntactic Structures*. Cambridge University Press, Cambridge.

Borer, Hagit. 2005. *The Normal Course of Events (Structuring Sense, vol. II)*. Oxford University Press, Oxford.

Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63 4: 741–782.

Cardinaletti, Anna and Michal Starke. 1999. The typology of structural deficiency: A case study of three classes of pronouns. In *Clitics in the Languages of Europe*, edited by Henk van Riemsdijk, pp. 145–233. Mouton, Berlin.

Chomsky, Noam. 1964. *Current Issues in Linguistic Theory*. Mouton, The Hague.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Ma.

Chomsky, Noam. 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. Harper & Row, New York.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English Transformational Grammar*, edited by Roderick A. Jacobs and Peter S. Rosenbaum, pp. 184–221. Georgetown University Press, Washington, D.C.

Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, edited by Stephen R. Anderson and Paul Kiparsky, pp. 232–286. Holt, New York.

Chomsky, Noam. 1977. On wh-movement. In *Formal syntax*, edited by Peter Culicover, Thomas Wasow, and Adrian Akmajian, pp. 71–132. Academic Press, New York.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11: 1–46.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. MIT Press, Cambridge, Ma.

Chomsky, Noam. 1986a. *Barriers*. MIT Press, Cambridge, Ma.

Chomsky, Noam. 1986b. *Knowledge of Language: Its Nature, Origins, and Use*. Praeger, New York.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, edited by Kenneth Hale and Samuel Jay Keyser, pp. 1–52. MIT Press, Cambridge, Ma.

Chomsky, Noam. 1995a. Bare phrase structure. In *Government and Binding Theory and the Minimalist Program: Principles and Parameters in Syntactic Theory*, edited by Gert Webelhuth, pp. 383–439. Blackwell, Oxford.

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

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by Step: Minimalist Essays in Honor of Howard Lasnik*, edited by Roger Martin, David Michaels, and Juan Uriagereka, pp. 89–155. MIT Press, Cambridge, Ma.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*, edited by Michael Kenstowicz, pp. 1–52. MIT Press, Cambridge, Ma.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and Beyond: The Cartography of Syntactic Structures, vol. 3*, edited by Adriana Belletti, pp. 104–131. Oxford University Press, New York.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36: 1–22.

Chomsky, Noam. 2007. Approaching UG from below. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*, edited by Uli Sauerland and Hans-Martin Gärtner, pp. 1–18. Mouton de Gruyter, Berlin.

Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, edited by Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta, pp. 133–166. MIT Press, Cambridge, Ma.

Chomsky, Noam. 2010. Some simple evo-devo theses: How true might they be for language? In *The Evolution of Language: Biolinguistic Perspectives*, edited by Richard K. Larson, Viviane Déprez, and Hiroko Yamakido, pp. 45–62. Cambridge University Press, Cambridge.

Chomsky, Noam. 2013a. The Dewey Lectures 2013: What kind of creatures are we? Lecture I: What is language? *The Journal of Philosophy* 110 12: 645–662.

Chomsky, Noam. 2013b. Problems of projection. *Lingua* 130: 33–49.

Chomsky, Noam. 2015. Problems of projection: Extensions. In *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*, edited by Elisa Di Domenico, Cornelia Hamann, and Simona Matteini, pp. 3–16. John Benjamins, Amsterdam.

Chomsky, Noam. 2017. The language capacity: Architecture and evolution. *Psychonomic Bulletin & Review* 24 1: 200–203.

Chomsky, Noam. 2019. Some puzzling foundational issues: The Reading program. *Catalan Journal of Linguistics* Special Issue 2019: 263–285.

Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. 2019. Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics* Special Issue 2019: 229–261.

Déchaine, Rose-Marie and Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33 3: 409–442.

Embick, David. 2010. *Localism versus Globalism in Morphology and Phonology*. MIT Press, Cambridge, Ma.

Emonds, Joseph. 2000. *Lexicon and Grammar: The English Syntacticon*. Mouton, Berlin.

Fitch, W., M.D. Hauser, and N. Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97 2: 179–210.

Frampton, John and Sam Gutmann. 2000. Agreement is feature sharing. Ms. Northeastern University.

Gärtner, Hans-Martin. 2002. *Generalized Transformations and Beyond: Reflections on Minimalist Syntax*. Akademie Verlag, Berlin.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40 3/4: 225–276.

Hauser, Marc D., N. Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–1579.

Hornstein, Norbert. 2009. *A Theory of Syntax: Basic Operations and UG*. Cambridge University Press, Cambridge.

Huang, C.-T. James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. thesis, MIT, Cambridge, Ma.

Lasnik, Howard and Juan Uriagereka. 2012. Structure. In *Handbook of the Philosophy of Science, Vol. 14: Philosophy of Linguistics*, edited by Ruth Kempson, Tim Fernando, and Nicholas Asher, pp. 33–61. Elsevier, Amsterdam.

May, Robert. 1985. *Logical Form: Its Structure and Derivation*. Linguistic Inquiry Monographs; 12. MIT Press, Cambridge, Ma.

McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. In *Derivation and Explanation in the Minimalist Program*, edited by Samuel David Epstein and Daniel T. Seely, pp. 184–226. Blackwell, Oxford.

Moro, Andrea. 2000. *Dynamic Antisymmetry*. MIT Press, Cambridge, Ma.

Muysken, Pieter and Henk van Riemsdijk. 1986. Projecting features and featuring projections. In *Features and Projections*, edited by Pieter Muysken and Henk van Riemsdijk, pp. 1–30. Foris, Dordrecht.

Nunes, Jairo. 1995. *Without Specifiers: Phrase Structure and Events*. Ph.D. thesis, University of Maryland, College Park, Md.

Ochi, Masao. 1999. *Constraints on Feature Checking*. Ph.D. thesis, UConn, Storrs, Ct.

Pesetsky, David. 2000. *Phrasal movement and its Kin*. Linguistic inquiry monographs; 37. MIT Press, Cambridge, Ma.

Pesetsky, David and Esther Torrego. 2001. T to C movement: Causes and consequences. In *Ken Hale: A Life in Language*, edited by Michael Kenstowicz, pp. 355–426. MIT Press, MIT, Cambridge, Ma.

Pesetsky, David and Esther Torrego. 2006. Probes, goals and syntactic categories. In *Proceedings of the 7th Annual Tokyo Conference on Psycholinguistics*, pp. 24–60. Hituzi Syobo, Tokyo.

Pesetsky, David and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation*, edited by Simin Karimi, Vida Samiian, and Wendy K. Wilkins, pp. 262–294. John Benjamins, Amsterdam.

Rezac, Milan. 2010. Φ-agree versus ϕ-feature movement: Evidence from floating quantifiers. *Linguistic Inquiry* 41 3: 496–508.

Rizzi, Luigi. 2015. Cartography, criteria, and labeling. In *Beyond Functional Sequence*, edited by Ur Shlonsky, vol. 10 of *The Cartography of Syntactic Structures*, pp. 314–338. Oxford University Press, Oxford.

Russell, Bertrand and Alfred North Whitehead. 1913. *Principia Mathematica*. Cambridge University Press, Cambridge.

Sells, Peter. 1985. *Lectures on Contemporary Syntactic Theories: An Introduction to Goverment-Binding Theory, Generalized Phrase Structure Grammar, and Lexical-Functional Grammar*. CSLI lecture notes; 3. CSLI, Stanford, Ca.

Sells, Peter. to appear. The view from declarative syntax. This volume.

Stabler, Edward. 1997. Derivational minimalism. In *Logical Aspects of Computational Linguistics*, edited by Christian Retoré, pp. 68–95. Springer, Dordrecht.

Starke, Michal. 2001. *Move Reduces to Merge: A Theory of Locality*. Ph.D. thesis, University of Geneva. Available at http://ling.auf.net/lingBuzz/000002.

Svenonius, Peter. 1994. C-selection as feature-checking. *Studia Linguistica* 48 2: 133–155.

Taraldsen, Knut Tarald. 1996. How to keep linguists happy: Review of the minimalist program by Noam Chomsky. *GLOT International* 2 3: 12–15.

Yang, Charles, Stephen Crain, Robert C Berwick, Noam Chomsky, and Johan J Bolhuis. 2017. The growth of language: Universal Grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews* 81: 103–119.

Zeijlstra, Hedde. 2008. On the syntactic flexibility of formal features. In *The Limits of Syntactic Variation*, edited by Theresa Biberauer, pp. 143–173. John Benjamins, Amsterdam.