Syntax for Cognitive Scientists

DRAFT, COMMENTS WELCOME

David Adger
Queen Mary, University of London

3rd September, 2013

1 Introduction

Syntax is the cognitive capacity of human beings that allows us to connect linguistic meaning with linguistic form. The study of syntax is a huge field that has generated a great deal of empirical and theoretical work over the decades (Everaert, Riemsdijk, Goedemans, and Hollebrandse 2005, Cinque and Kayne 2008, den Dikken 2013). In this article, I outline why understanding our syntactic capacity is important to cognitive science in general and why the data of syntactic research is to be taken seriously. I then sketch a number of very broad findings about the character of the structures of human language, keeping an eye on evidence from both syntactic and psychological/neuropsychological research. I conclude with a brief discussion of theoretical models.

We focus on syntax here, not language in general. Researchers outside of linguistics typically take the term language to have a general common-sense meaning, connecting it to its functions in communication, sociality, creativity, cultural transmission, thinking etc. While these functions are fascinating and important, I take it to be possible to investigate the syntax of language while leaving questions of its functions aside. Without an understanding of the nature of the capacity, questions as to how (or whether) its function has shaped that nature are shots in the dark.

2 Why is syntactic theory relevant to cognitive science?

Sentences (and other expressions) of human languages have structure. They are not simply sets of words or word-parts. This structure is relevant to the meaning of sentences and to their pronunciation. It is why *Anson bit Lilly* does not mean the same as *Lilly bit Anson*, even though both expressions consist of the same words. It is also why the words *her* and *she* differ in pronunciation, even though they refer to the same individual, and play a close to identical semantic role, in sentences like *Anson believes her to be vicious* and *Anson believes that she is vicious*.

One might think that, in the cases just mentioned, what is relevant is merely the order or perhaps the local context of the words, and that that might be sufficient to determine both meaning and form. To see that simple order will not suffice, compare *Anson bit Lilly* with *Lilly was bitten*. Here the word *Lilly* precedes rather than follows *bite*, but it has the same meaning with respect to *bite* (that is, the same semantic role) as in *Anson bit Lilly*: in both cases, Lilly suffers the bite. At a slightly more abstract level, take the sentences *Lilly fears Jasper* and *Lilly amuses Jasper*. In the first, the individual who is feeling the emotion is denoted by the word that precedes the verb, while in the latter, the individual feeling the emotion follows the verb. Such effects, where meaning is divorced from word order, are ubiquitous in language.

The point here is not only a descriptive one. It would be trivial to define a learning algorithm for a language that defined semantic roles on the basis of linear order (like, for example, standard translations of English into predicate logic do). The algorithm would simply seek to match cognitive semantic concepts relevant to meaning to surface properties of the ordering of words (or rather categories of words); however, no human language works like this.

The argument that structure trumps not only linear order but also local context is equally straightforward and is built on the existence of dependencies that hold between items that are linearly distant from each other between with sentences.

For example, in the general case, the form of a verb is not determined by linearly adjacent words. This is why in *The girls from Paris are singing*, the auxiliary verb form of *be* appears as the plural version *are* and not the singular *is*, even though *are* is directly adjacent to the singular noun *Paris* but distant from the plural noun *girls*. Compare this with *Paris is beautiful*, which shows that *Paris* can trigger a singular form of the verb. The examples differ because *Paris* is in a different structural relationship to the verb in the two cases, even though it provides the same local context for the verb.

Such simple examples are important because they are general across human languages and reveal an important feature of how language works: no language that we know of has a general rule that will trigger verbal agreement in number with the equivalent of *Paris* in *The girls from Paris are singing*; that is, linear contiguity is never the precondition for this kind of verb agreement. Even more interesting is that speech errors often give rise to such agreement ('attraction errors' Bock, Eberhard, Cutting, Meyer, and Schriefers 2001), but these are never generalized by children learning the language to a rule that causes the verb to agree with the adjacent noun. If they were so generalized then the result would be a paradigm that looks as follows, with the reverse of the usual English judgments in (a-d) but the standard English judgments in (e-h) (I present data from now on numbered as is standard in linguistics, with a * to signify an empirical claim that the sentence is not acceptable in the language in question. Section 3 discusses the nature of syntactic data in more depth):

(1) a. The girls from Paris is singing
b. *The girls from Paris are singing
c. The girl from the Western Isles are singing
d. *The girl from the Western Isles is singing
e. Paris is beautiful
f. *Paris are beautiful
g. *The Western Isles is beautiful
h. The Western Isles are beautiful

Not only does no variety of English work like this, no variety of any human language works like this. From the perspective of cognitive science, this is fundamental, as it is again trivial to define a learning algorithm that will look at the preceding noun in a linear sequence and determine the agreement on the following verb on the basis of that noun's properties. A simple Markov model providing a local linear context would suffice and, given that agreement is particularly vulnerable in language change (Ferguson 1996), and attraction errors exist (Bock, Eberhard, Cutting, Meyer, and Schriefers 2001), the prediction would be that languages would drift towards a linearly based agreement system. However, they do not. Rather human language learners ignore the linear adjacency relationship as a potential hypothesis and successfully learn that even distant nouns can be appropriate governors of the verb form:

(2) The **girls** from Paris that the cat scratched in the house **are** singing.

This entails that the hypothesis space that human language learners search in attempting to learn how meaning relates to form in their language is restricted so that it includes hypotheses involving structural conditions on dependencies and excludes those involving linear conditions (Chomsky 1968, Chomsky 2013).

The problem of determining the verb form in the examples here is solved by taking *girls* to have a structural, rather than a linear, relationship, to the verb (equivalently, what is relevant is structural context, not linear context). Most current theories of syntax adopt the view that there is a structural unit superordinate to *girls* but sharing properties with it, and that unit is in a local relationship to the verb. Schematically, this looks as follows:

(3) $[_{\text{plural}}$ the $\text{girls}_{\text{plural}}$ from Paris that the cat scratched in the house] $\text{are}_{\text{plural}}$*singing*

The structural distance between the plural property of this unit (marked by the square brackets) and the plural property of the verb is then minimal, and this minimal distance also allows the semantic relationship between the two to be determined. In this way there is indeed a contiguity between the verb and the expression that governs its form, but the contiguity is of two structural not linear units. The plural property of the whole unit is inherited, again in a structural fashion, from an element inside it. This works by assigning a structural relationship between *girls* and the whole superordinate unit (a relationship that is not shared by *Paris* or *cat* and that unit). Other analyses have been proposed, but all rely on structural not linear relations.

What of the semantic relationship between *the girls* and *scratched* in (3)? Somehow *the girls* is interpreted as the individuals who are scratched. Most current theories of syntax agree on the core properties of this phenomenon although they differ in how they encode these: some propose that the relevant property of the verb (that it may combine with a grammatical object that is interpreted as the entity that gets scratched) is realised as an abstract property of the structure and this is then percolated through the structure until it can be identified by *girls*; others propose that the construction of the superordinate unit involves a simultaneous abstract contiguous relationship between *girls* and *scratched* and between *girls* and the determiner *the*, so that the single element *girls* bears two structural relationships: one to *scratched* and one to *the*. Schematically, these two options look as follows (where italicization of *girls* marks that it is a single element appearing in two parts of the structure simultaneously):

(4) $[_{\text{plural}}$ the $\text{girls}_{\text{scratchee}-\text{object}}$ from Paris that the cat $\text{scratched}_{\text{scratchee}-\text{object}}]$ $\text{are}_{\text{plural}}$ singing

(5) $[_{\text{plural}}$ the *girls* from Paris that the cat scratched *girls*$]$ $\text{are}_{\text{plural}}$ singing

We will return to these issues in section 4.4, but what is crucial for present purposes is that all theories of syntax agree that a dependency based on abstract structure has to be established between *girls* and *scratched* in cases like this, though the technical means of doing this varies from theory to theory. The non-contiguity of both form and meaning relationships in language is a straightforward discovery of syntactic investigation with consequences for models of cognition more generally.

The proposal that structural not linear properties are fundamental to human language, stable since the 1950s in linguistics, has a great deal of psychological and neuropsychological confirmation. Neuroscientfic investigation has consistently found that there are different patterns of brain activation when humans learn artificial languages which are based on structure vs linearity. For example, Musso, Moro, Glauche, Rijntjes, Reichenbach, Büchel, and Weiller (2003) showed that subjects' Broca's areas were activated in an increasing fashion over time for structure based learning as opposed to linear based learning, while Frederici, Bahlmann, Heim, Schubotz, and Anwander (2006) showed evidence for cortical separation of hierarchical from linear processing. See Berwick, Friederici, Chomsky, and Bolhuis (2013) and especially Friederici (2011) for a comprehensive review. We return to evidence for hierarchical structure in section 4.1.

Syntax, as a discipline, focuses on the question of the character of the structures of human language and provides (part of) the account of what deeper principles lead them to be that way. There are therefore two distinct tasks in the field: careful analytical development of proposals for the description of the syntactic properties of various linguistic phenomena both within and across languages, and theoretical development of a model which embodies the stable abstract generalizations that emerge from this work. There is a good deal of consensus about analysis, with various insights about structure developed in different theoretical frameworks. There is less consensus about the appropriate theoretical model, although even here there is more than may appear to those outside the field (see section 5).

3 The question of data

Psychologists and syntacticians work on very different kinds of data. For the most part, the data that syntacticians work with will have medium to large effect sizes (Cohen's *d*, Cohen 1988): that is, the mean difference in a measured behavioural reaction to a stimulus divided by the mean standard deviation is much greater than 0.5. Effect size is a measure of the strength of a phenomenon and, as noted by Sprouse and Almeida (2013), Cohen defines medium to large effect sizes as those that a trained researcher can see without applying statistical analysis. Sprouse and Almeida in a series of papers summarized in Sprouse and Almeida (2013) show that, for a large and representative sample of syntactic data (collected via traditional, informal methods), formal experimental investigation converges with the informally collected data (97% of the relevant phenomena covered in a graduate textbook (Adger 2003) and 95% of the English data in a decade of Linguistic Inquiry articles). They attribute this in part to the fact that the data used to make scientific claims about the syntax of human languages, in general, have medium to large effect sizes.

To get a sense of the issues, take a simple set of sentences like the following:

(6) a. I met someone from New York at the conference.
    b. From New York I met someone at the conference.
    c. At the conference I met someone from New York.

The empirical task the syntactician sets him/herself is to determine whether (b) or (c) is most closely related to (a). Without performing any statistical test, it's clear that (c) is related to (a) more closely than (b) is, and, in fact, (b) is not acceptable as a sentence of English with the same meaning as (a), while (c) is. This is a typical datum in syntax, and in fact it can be used as an argument in a fairly sophisticated chain of reasoning about what the structure of (a) is. The details of the analysis aren't relevant here, but the strength of the effect size is evident to any speaker of English.

This kind of data is exactly the kind of data that most syntacticians work with on a daily basis, and the strength of the effect is so clear that most syntacticians feel justified in not subjecting it to statistical testing (for discussion of this issue see Culicover and Jackendoff 2010, Schütze and Sprouse 2013, Sprouse and Almeida 2013, but see Gibson and Fedorenko 2010 for a dissenting view). Further, the speed with which this kind of data can be amassed has allowed syntacticians to build a large and solid empirical understanding of syntax across many languages. For example, translating the sentences above into other languages, perhaps languages with quite different grammatical properties from English, can quickly lead to an understanding of how consistent or variable the phenomenon is, and to what extent it correlates with various other properties.

Does this mean that syntacticians should not perform experiments to test the strength of their claims? Certainly not. In some cases the effect size is not clear to the researcher. Until recently, the tack taken in the discipline has been somewhat Darwinian: the field will converge on those datapoints where researchers agree (through article reviews, conference presentations, discussions with colleagues, supervisors etc.) that the effect size is large enough to be detected without statistical testing. Sometimes this will involve more careful design of the materials to weed out pragmatic or semantic factors that interfere; sometimes data will simply not be considered to have a strong enough effect size to be used in theoretical or analytical argumentation. More recently, however, experimental work in syntax has taken off (Myers 2009, Gibson, Piantadosi, and Fedorenko 2011, Erlewine and Kotek 2013) and for cases where the effect size is not large enough but the data is important for casting light on some theoretical issue, experimental evidence becomes an important input to the theoretical debates.

The lack of statistical and other information about data provenance and analysis does raise a communication difficulty: psychologists looking at syntax papers and seeking the kind of statistical information they are used to in their own discipline will generally find it lacking. However, the field itself clearly has mechanisms for ensuring the solidity of the data that it uses to build theoretical claims on, and as Sprouse and Almeida's work shows, the vast majority of the data used in theoretical argumentation are replicable and reliable, because the size of the effect is immediately and easily checked. Different disciplines have different methods of establishing their core phenomena. Judgments of acceptability applied to well designed materials in syntax have been crucial in building a very rich base of cross-linguistically valid knowledge about syntactic structure.

4 Important findings in syntax

In section 1 we saw that structure, not linear order, is relevant to how meaning and form are related in human language. An equally important empirical point about the meaning-form relationship is that it is astoundingly vast: a speaker of a language, who has acquired a finite set of words in that langauge, can link the forms of sentences in the language to their meanings over a range of experiences so large that it seems senseless to place an arbitrary limit on it. Cortical and psychological evidence suggests that learning the grammar of a language is complete by puberty (Kim, Relkin, Lee, and Hirsch 1997, Newport 1990). It follows that speakers acquire a productive means to link form and meaning (a grammar) over a practically unbounded range, from a bounded set of experiences. The question is how to model such a capacity.

Perhaps the major insight of generative syntax is that a human grammar can be modelled by a finite mathematical function. This function, in the early years of investigation, was taken to be fairly complex, modelling a number of levels of linguistic description, and leading to a characterization of human linguistic knowledge that seemed highly particular (e.g. the models in Chomsky 1965, Chomsky 1976, Chomsky 1981). The general framework, in its many incarnations, allowed an explosive growth in empirical knowledge about the syntax of languages as well as increasing depth of analysis and a series of well understood high-level generalizations about phenomena. More recent models have radically reduced in complexity as the empirical generalizations have become clearer (Rizzi 2013 for brief review), allowing far less innate structure to be imputed to the linguistic system (Hauser, Chomsky, and Fitch 2002).

We can divide the high level generalizations discovered over the years into two broad classes: generalizations to do with the shape and interpretation of structures and generalizations about dependencies between elements within these structures.

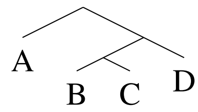
4.1 The reality of constituent structure

The notion that sentences are made up of linear sequences of smaller components was already challenged in section 1. All approaches to syntax agree that such a view is empirically inadequate and that some means is necessary to represent non-linear properties of linguistic expressions. One such property is constituent structure.

Constituent structure is simply the grouping of linguistic elements together to the exclusion of other elements, establishing another dimension of organization than the linear order of the elements in a sentence. Usually this is represented as either a tree-like structure, or a bracketting:

(7) a. [A [[B C] D]]

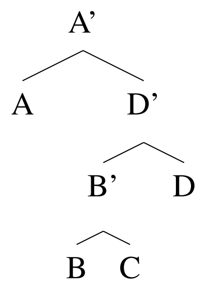
b.



The expressions grouped into a single constituent are said to be sisters, while the overarching constituent that contains them is termed the mother. Typically, it is assumed that there is information specified at the juncture points in the tree (the non-terminal nodes) as well as at the terminal nodes of the tree. This means that these nodes are labelled or decorated in some fashion. For example:

(8) a. $[_{A'}$ A $[_{D'}$ $[_{B'}$ B C] D]]

b.



Labelling is distinct from constituency, and indeed some approaches in syntax eschew it (Collins 2002) preferring to specify all the relevant information in the terminal nodes.

Evidence for constituent structure comes from a range of linguistic domains.

(i) Constituents allow a description of distributional patterns in languages that allows linguists to capture a wide range of facts about particular languages via a condensed set of constituent structure types; this is what allows us to capture the fact that a pronoun like *she* and a phrase like *David and Anson's sleepy cat* are intersubstitutable in English in a wide range of syntactic contexts (they are all characterizable as a constituent with the same kind of node label), while at the same time capturing the fact that there is no lexical item that is intersubstitutable with, say, *and Anson's sleepy* which is a subsequence of *David and Anson's sleepy cat* but not a constituent of it. Such distributional generalizations were the initial motivation for constituent structure (Wells 1947).

(ii) Many overarching syntactic generalizations that govern sentence structure are typically statable in terms of constituents. For example, in many Germanic languages, such as Swedish, the verb in a main clause must be preceded by exactly one constituent (Holmberg 2013, examples 1-3):[1]

(9) a. I **have** honestly speaking never seen adders in this here forest

b. Adders **have** I honestly speaking never seen in this here forest

[1]When I present material from languages other than English, I will usually provide glosses expressing order, meaning and grammatically relevant information, rather than the data itself, on the assumption that most readers will be interested in the patterns ratrher than the specifics of the data. For those interested, the references provide the source and example numbers of the data.

c. In this here forest **have** I honestly speaking never seen adders

d. Honestly speaking **have** I never seen adders in this here forest

The cases in (9) are representative examples of a phenomenon which is best characterized by requiring that the tense marking auxiliary verb (**have**) appears after the first constituent of the sentence.

More generally, a huge range of syntactic generalizations are sensitive to constituents (see any syntax textbook for a discussion Haegeman 1991, Adger 2003, Carnie 2010). The generalizations are established within languages and typically converge on identical assignments of constituent structure (though see Pesetsky 1995, Phillips 1996, Steedman 1990 for cases where there are interesting mismatches), and the methods for determining constituency are applicable across languages.

(iii) Constituents display both semantic and phonological unity to the extent that theories of sentence semantics (Heim and Kratzer 1998) and phonology (Selkirk 2011) make crucial use of constituent structure in explaining how meaning is composed, prosody assigned and certain across-word phonological processes determined (although there are theories that avoid constituents that do this, e.g. Steedman 1996). Relatedly, assigning independently motivated constituent structures correctly predicts a range of semantic ambiguities without the need for special stipulations in the semantic component of the grammar. For example, it is possible to modify a noun with an adjective as in *sleepy cats* and it is also possible to combine two nouns into a single constituent as in *cats and dogs*. It follows with no further stitulation that the phrase *sleepy cats and dogs* will be ambiguous, since that sequence of words will be consistent with two assignments of constituent structure (I assume a binary constituency here, although that is not relevant to the argument):

(10)

In the structure on the left the adjective *sleepy* modifies the whole coordinate structure *cats and dogs* while in the tree on the right it modifies just *cats*, correctly capturing the ambiguity (and, incidentally, the natural position to place pauses if a speaker wishes to prosodically disambiguate the phrase). It also follows that *cats and sleepy dogs* will not be ambiguous, as only one assignment of constituent structure is available. These kinds of phenomena are pervasive and constituency provides an elegant way of capturing them.

(iv) Psychological evidence abounds that these constituent structures (or at least some of them) are relevant in sentence processing: the structural priming phenomenon (Bock 1986, Pickering and Ferreira 2008 for review), perception experiments (Fillenbaum 1971; Fodor, Bever, and Garrett 1974, Cohen and Mehler 1996), relatedness judgments (Levelt 1970), slips of the tongue effects

(Fromkin 1971) etc.

While most approaches to syntax involve the specification of constituent structure, some, including approaches in categorial grammar (e.g. Steedman and Baldridge 2011) and dependency grammar (e.g. Hudson 2010) derive the same clustering effects in different ways.

4.2 Further properties of constituent structure

Early approaches to the modeling of constituent structure were extremely powerful. The posited rule systems, for example, defined constituent structures in a way that, in principle, allowed a larger unit to be built out of any number of any types of smaller units in any order. For example, early models such as the Phrase Structure Rule model would allow a rule set like the following (Lyons 1968):

(11) a. VP $\rightarrow$ N P A N
b. VP $\rightarrow$ N N A P

Here a verbal phrase (notated VP for V(erb) P(hrase)) is defined to be made up out of a noun (N), a preposition (P) (itself consisting of just a verb) an adjective (A) and a noun, with two possible orders for the preposition and the final noun, allowing them to be swapped around. However, cross-linguistic empirical work made it clear that the number of branches that could subtend from a single node (its daughters) was limited, that the type (or category) of a node was not random, but was predictable from its daughters, and that generalizations about order could be extracted out of the constituent structure system, divorcing the linear sequence of nodes from their hierarchical organization. These were empirical findings that motivated a series of constraints intended to more accurately characterize human grammars.

Current views of syntax take the number of subunits of a constituent to be severely limited. This makes sense from a learnability perspective. Take, for example, the proposal that syntactic constituency is limited to two subconstituents per unit (that is, thought of as a tree structure, all nodes would branch in a binary fashion, Kayne 1984). If we have a string of say four words, then in the absence of any constraint on branchingness, there will be 11 ways of bracketing this:

(12) a. [might have been running]
b. [[might have been] running]
c. [might [have been running]]
d. [might [have been] running]
e. [[might have] been running]
f. [might have [been running]]
g. [[might have] [been running]]

h. [might [[have been] running]]
i. [might [have [been running]]]
j. [[might [have been]] running]
k. [[[might have] been] running]

If, however, branching is constrained to be binary, the number reduces by more than half (generally, for a string of n words, the number of binary branching structures is the Catalan number of n-1):

(13) a. [[might have] [been running]]
b. [might [[have been] running]]
c. [might [have [been running]]]
d. [[might [have been]] running]
e. [[[might have] been] running]

If it turns out that the syntactic structures of human language are binary in nature, we can immediately simplify the task of the learner.

Linguistic evidence that, in fact, for the case of English auxiliary verbs just presented (one of) the binary structures is correct (13-c) has been known since Ross 1969. The notion that branchingness is limited is the consensus position in the field, although whether strict binary branching for all structures is adopted varies from model to model.

It is also a close to consensus position that linear order is to be factored out from a specification of hierarchical structure (Stowell 1981; Gazdar, Klein, Pullum, and Sag 1985, but see Kayne 1994 for a dissenting view). That is, the grammatical system specifies constituent structure independently of the order of the elements in that structure. So the two trees in (14) are equivalent ways of representing the same object:

(14)

The strongest argument for this position comes from the elegant analysis it gives of cross-linguistic generalizations that keep hierarchy separate from constituency. Consider the English sentence:

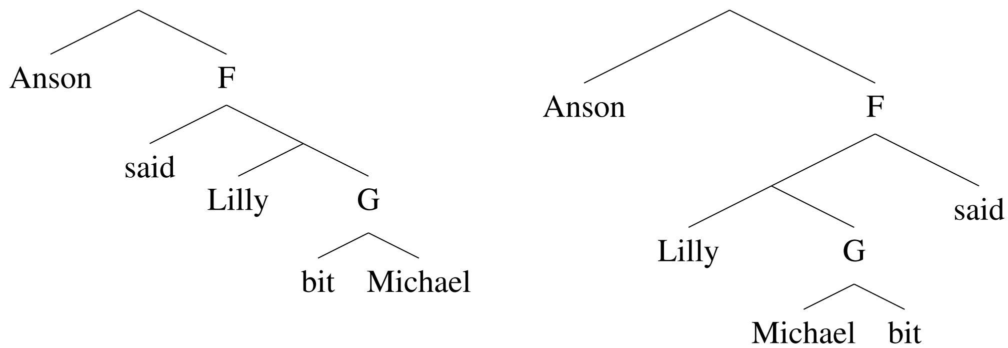
(15) Anson said Lilly bit Michael

In Japanese, this translates to the equivalent of (16) (I leave out representations of Japanese case markers and complementizers here, as they are irrelevant to the point):

(16) Anson Lilly Michael bit said

For both languages, the language specific evidence is that the constituency is identical (Adger 2003, Tsujimura 2013): The verb *bite* forms a constituent with *Michael* and this unit forms a constituent with *Lilly*. That whole unit forms a constituent with the verb *say* which then forms a constituent with *Anson*. We thus have the following where the structure is exactly the same, but the order is reversed under the nodes F and G:

(17)



In fact, binary branching and allowing either order of sisters together radically restrict the number of possible structures that a learner of a language can posit. Take again our simple sentence *Anson said Lilly bit Michael* which has five words. That means that there are 5!=120 possible permutations of this string. Now let's impose a (correct) binary branching constituency on these and assume each binary branching constituent has no specified order. This results in just 15 possible orders:

(18)
a. [Anson [said [Lilly [bit Michael]]]]
b. [Anson [said [Lilly [Michael bit]]]]
c. [Anson [[Lilly [Michael bit]] said]]
d. [Anson [[Lilly [bit Michael]] said]]
e. [Anson [said [[Michael bit] Lilly]]]
f. [Anson [said [[bit Michael] Lilly]]]
g. ...

This idea immediately rules out a number of cases that are unattested in human language. For example, we know of no language where reversing the two verbs in (15) is a way of expressing (15):

(19) *Anson bit Lilly said Michael (with the meaning *Anson said Lilly bit Michael*)

It also neatly captures some common kinds of languages. While (a) in (18) is English, (b) is the common surface order of other Germanic languages, (c) covers languages like Japanese while (d) is a possibility in Basque (although not the only order). Further these are roughly the independently motivated constituency structures for these languages.

The orders from (e.) onwards are controversial: they involve taking the constituent which is the sister of the verb phrase (which we can call the *subject* of the verb phrase) and placing it to the right of that verb phrase. At a rough estimate, less than 2% of the world's languages are subject final in their surface order (The World Atlas of Language Structures gives gives 36 languages out of 1377). Most syntacticians working on languages which place the subject to the right do not analyse this as an effect of the ordering of the subject and its sister verb phrase, but rather take it to involve a non-transparent mapping between hierarchical structure and linear order (Chung 2005). The contrast in the frequency of languages that can be analysed by placing the sister of the verb (that is, the verb's object) on either side of the verb, versus placing the sister of the verb phrase (that is, the verb phrase's subject) on either side of the verb phrase is striking. If subjects cannot be placed rightwards by a simple ordering of sisters, the structures in (e) onwards are never possible hypotheses for a language learner, further restricting the range of analyses open to the learner (but requiring a different analysis of subject final orders). If subjects can be ordered to the right of their sisters, then an alternative explanation is required of the language internal data that has been used to argue against this analysis.

However, the various orderings in (18) are actually not sufficient to analyse the basic orders of constituents in human languages. One major set of structures not covered is represented by the following order:

(20) Said Anson bit Lilly Michael

Here the verb is placed before the subject (unlike in English where it follows the subject). These Verb-Subject-Object (VSO) languages are about 15% of the WALS sample and include various languages from Celtic, Semitic, Austronesian, Mayan, Afro-Asiatic and elsewhere). Simple reorderings of sisters in binary constituents are not sufficient to generate or analyse such structures. There are two obvious approaches: one takes VSO to be a flat ternary structure which makes the subject a sister of the verb and hence freely orderable under sisterhood, another takes VSO to involve a different kind of permutation from free ordering within a binary constituent (Emonds 1978). The former proposal faces many problems of analysis when applied to actual languages (Anderson and Chung 1977, Sproat 1985), and is generally not accepted. The alternative posits a permutation that derives the VSO order from a structure where the verb and object form a constituent by placing the verb to the left of the subject, a structure for which there is much constituency evidence (McCloskey 1983):

(21) Subject [Verb Object] $\rightarrow$ Verb [Subject [Object]]

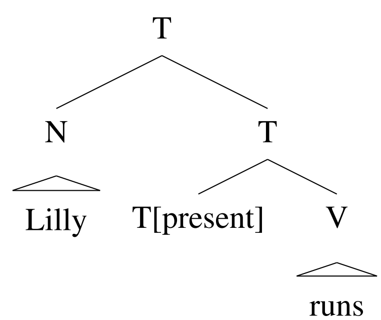
(22) Said [Anson [bit [Lilly Michael]]]

Allowing such a permutation will derive the VSO order from any of the allowed structures in (18): all of them will dervive the same surface form. This does however raise the question of how to model such permutations in a restrictive way. [2]

A final comment about languages with free word order is necessary. Even those languages which have been claimed to have extremely free word order (for example the Pama-Nguyan language Warlpiri or the Kiowa-Tanoan language Kiowa), where there seems to be no obvious evidence for constituent structure of the sort discussed so far, have turned out not to involve random permutations on a string or very flat structure. Contrary to the original proposal of Hale (1983), detailed empirical work has shown that these languages have clear hierarchy effects, suggesting constituent structure (Legate 2002, Adger, Harbour, and Watkins 2009). However the debate is still open as to how to model these empirical findings.

We have seen that branching is likely restricted and that order is disconnected from hierarchy. The third important consensus position on properties of constituent structure is the idea that the type or category of a constituent is predictable from the type or category of a constituent contained within it (the head). This is known as endocentricity, or headedness, and it further constrains the range of analyses available to a learner. The constraint on headedness incorporates the generalization that the syntactic distribution of a constituent is closely related to the syntactic distribution of one of the constituents it immediately contains. This constraint rules out many classical approaches to clause structure that take a sentence to be a distinct kind of grammatical entity from the noun (phrase) subject and verb phrase it contains. Endocentricity forces a model of the sentence where it is either headed by NP or VP, or where it has a distinct abstract head. The latter view is now the consensus position, with the category sentence itself being more layered than in the classical analysis: for English, at least, it has as its head a category that semantically specifies temporal information (notated T) so that a simple sentence like *Lilly runs* looks as follows, where T is the label of those nodes of which the element specifying tense information is the head.

(23)



[2] There is an open question about whether this operation is one that applies directly to constituent structure representations, changing the hierarchical order, followed by linearization of the tree (so called head-movement analyses Koopman 1984, Roberts 2005), or whether it is an operation that is part of the linearization of the tree (direct linearization analyses Brody 2000, Adger 2013).

Categories like T are known as functional categories, and have become extremely important in modern syntactic theory (Muysken 2011). Functional categories are the closed-class categories of traditional grammar, and there is agreement that they are linguistically distinct from the categories that distinguish content words. In the latter case we have, for example, *cat* being an N, *jump* a V and *happy* an A(djective), where the lexical categories N, V and A distinguish classes of content words, each class having many members. Functional categories label far fewer expressions (e.g. the (C)omplementizer category introduced below has only three members in English) which means that particular tokens of functional categories are frequent in contrast to tokens of lexical categories with concommitent implications for acquisition (Hochamnn 2013). In addition, there are a number of linguistic and psychological differences. Phonologically, the expressions labelled by functional categories tend to be shorter, non-stress bearing and often have particular sets of phonological constraints imposed upon them (such as a restricted set of consonants or vowels, Muysken 2009). They are also often affixes, attaching to contentive categories. Semantically, they specify grammatically relevant meanings such as tense, aspect, definiteness, plurality, etc. These are rigid, fixed points in meaning (Keenan and Stabler 2004), which can be specified using well-understood logical techniques. In this they contrast with the meanings of content words like *run*, which are semantically variable and resist complete analysis using logical techniques.

Syntactically, functional categories provide positions in constituent structure, thus taking over the role of phrase structure rules. Thus in the tree structure in (23), the category T combines with the verbal expression *runs*, and the result combines with the subject noun *Lilly*, so that T serves as a kind of syntactic glue, sealing together the subject and the verb phrase, while semantically expressing the temporal constitution of the event of running (in this case that it is a habitual activity of Lilly's). Expanding beyond these domains, functional categories and contentive categories are differently affected in language disorders (Penke 2011), language acquisition (Gervain and Mehler 2011, and neurophysiological response (Diaz and McCarthy 2009).

The presence of functional categories, plus the constraint on binary branching, leads to a fair amount of depth in syntactic representations: it is not untypical to see tree like representations spanning whole pages of articles (or for syntax teachers to run out of whiteboard). However, this depth should not be confused with abstractness. In a fully fledged analysis, each functional category is endowed with a semantic interpretation, so each is semantically concrete. Most have either direct phonological expression in the relevant syntactic position, or are in a syntactic dependency with another element that phonologically expresses the particular semantic value of the category. For example, in the tree above, T marks the temporal semantics, but the overt expression of T appears as an inflection on the verb, possibly non-contiguous with T (as in *Lilly often runs*). Contrast *Lilly runs* with the emphatic structure *Lilly does run*, where the inflection appears not on the verb, but on an auxiliary preceding the verb (that auxiliary is taken to be placed in the T position). We take up the nature of such dependencies below.

The TP category is in constituency with another functional category, the complementizer (C). C may be overtly expressed in English when a sentence combines with a verb:

(24) a. I know **that** Lilly runs.
b. I asked **if** Lilly runs.

The semantics of whether the embedded sentence *Lilly runs* expresses a fact or asks a question is marked overtly by the distinction between the complementizers *if* and *that*. We might ask whether a non-embedded (matrix) sentence also has a C in English. The answer appears to be that it does:

(25) Does Lilly run

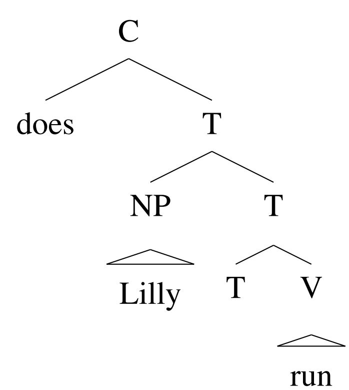
Here the matrix sentence is asking a question, and the overt marking of that question is the appearance of the auxiliary verb *does* before *Lilly*. In other languages, such as Celtic language, the question C in a matrix clause is simply identical to the question C in an embedded clause, so that, in Scottish Gaelic, for example, the way to ask (25) is the equivalent of *If Lilly runs*. For English, the null hypothesis would be that sentences always have a C, and that in (25), C is filled by *does*, just as it is filled by *if* in (24-b). Corroborating evidence for this comes from varieties of English where examples like (25) can appear as embedded sentences, for example certain varieties of Irish English where both (26-a) and (26-b) are well formed sentences:

(26) a. I asked does Lilly run. Irish English Henry (1995)
b. I asked if Lilly runs. Irish English

In such varieties, however, it is impossible to have both *does* and *if*, suggesting that there is a single C position:

(27) *I asked if does Lilly run.

(28)



Many functional categories are strictly hierarchically ordered with respect to each other. For example, taking *that* to be a C, and emphatic *does* to be T, we find only one order is possible, captured by requiring C to be hierarchically superior to T:

(29) a. I know that Lilly does run
b. *I know does Lilly that run.

This kind of argumentation about distribution, semantics, constituency and order has been used to establish a number of functional categories that appear in both sentences and noun phrases (and sees its apogee in the Cartographic approach to syntax (Cinque 1999, Cinque and Rizzi 2010)). For example, in the noun phrase in English we find that a demonstrative (Dem) like *this/these* occurs before a numeral (Num) like *six*, which in turn occurs before an adjective like *interesting*:

(30) a. These six interesting ideas
b. *Interesting six these ideas

Recall that linear order is a side effect of hierarchical structure; standard constituency diagnostics suggest the following structure for such phrases (Adger 2013), where each 'layer' in the structure is labelled with the relevant functional category:

(31) $[_{\text{Dem}}$ These $[_{\text{Num}}$ six [interesting ideas]]]

If we allow the binary branches to permute, we have a further seven possible orders:

(32) a. $[_{\text{Dem}}$ These $[_{\text{Num}}$ six $[_{\text{A}}$ ideas interesting]]]
b. $[_{\text{Dem}}$ These $[_{\text{Num}}$ $[_{\text{A}}$ interesting ideas] six]]
c. $[_{\text{Dem}}$ These $[_{\text{Num}}$ $[_{\text{A}}$ ideas interesting] six]]
d. $[_{\text{Dem}}$ $[_{\text{Num}}$ Six $[_{\text{A}}$ ideas interesting]] these]
e. $[_{\text{Dem}}$ $[_{\text{Num}}$ Six $[_{\text{A}}$ interesting ideas]] these]
f. $[_{\text{Dem}}$ $[_{\text{Num}}$ $[_{\text{A}}$ Interesting ideas] six] these]
g. $[_{\text{Dem}}$ $[_{\text{Num}}$ $[_{\text{A}}$ Ideas interesting] six] these]

In fact, all of these orders are attested cross-linguistically (see Cinque 2005 for details) but in addition we find orders where the noun appears to the left, parallel to what we saw with verb subject object (VSO) ordering in the clause:

(33) a. Ideas $[_{\text{Dem}}$ these $[_{\text{Num}}$ six $[_{\text{A}}$ interesting]]]
b. Ideas $[_{\text{Dem}}$ $[_{\text{Num}}$ six $[_{\text{A}}$ interesting]] these]
c. Ideas $[_{\text{Dem}}$ $[_{\text{Num}}$ $[_{\text{A}}$ interesting] six] these]
d. Ideas $[_{\text{Dem}}$ these $[_{\text{Num}}$ $[_{\text{A}}$ interesting] six]]

Further, it has been known since Greenberg (1963) that there are no languages where the noun is final but the various other elements in the sentence permute (Greenberg's Universal 20):

(34) *Interesting six these ideas.

This range of typological variation is captured by a system of binary branching, free permutation, and displacement of the noun to the left, if we take the hierarchy of these elements to be universally fixed (Abels and Neeleman 2012, or Cinque 2005 for the same result with a slightly different model). If there is no displacement of the noun so that it ends up linearised to the right of the rest of the structure, the typological gap in (34) emerges, and the structural analysis of the noun phrase and the sentence come into line.

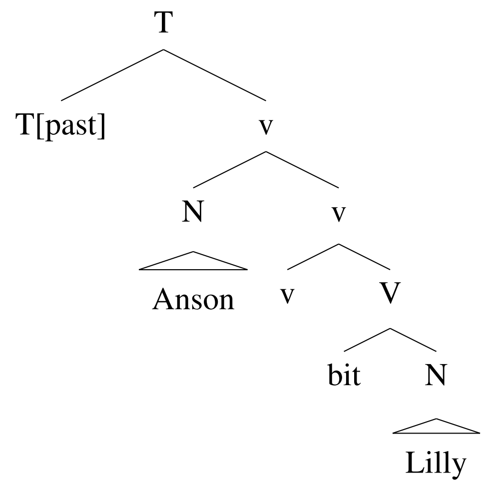
The kind of approach outlined here takes the basic categories of noun and verb to be surmounted by a (hierarchically) ordered sequence of functional categories carrying logical semantic information. Such sequences are termed Extended Projections (Grimshaw 1991). There is a great deal of debate about the richness, hierarchical order and universality of these functional categories, but there is a strong consensus that they are required.

This section has sketched some important lines of research that limit the class of attainable grammars via constraints on branchingness, on the linear permutations of hierarchical structures, on the labelling of hierarchical structures via functional categories expressing particular kinds of semantic information. These various constraints not only provide good explanations of syntactic phenomena within languages, but also of typological observations across languages.

4.3 The Interpretation of Structure

We have already touched upon the interpretation of functional categories. These add extra semantic information to the basic predicate or concept introduced by the noun, verb, adjective or other content word. We now turn to the interpretation of the contentful categories, restricting our attention to verbs. Verbs typically occur with noun phrases that are semantically connected to them: for example, a verb like *jump* has a single participant, expressed by the noun phrase *the cat*, in *the cat jumped*. A number of important generalizations have been discovered about these verb-noun phrase relations over the years. We will discuss just one: the noun phrase that denotes the agent of an event is hierarchically superior to the noun phrase that denotes a non-agent (Jackendoff 1972 and much work following this). In the example, *Anson bit Lilly*, the subject *Anson* is external to the constituent *bit Lilly*, hence hierarchically superior to *Lilly*, and it is interpreted as the agent of the event. This correlation between semantic roles and syntactic hierarchy is very strong, and the rare apparent exceptions reduce to independent factors. One way that this generalization has been embedded in theoretical models is to take the subject to be introduced by a functional category whose meaning links the meaning of the subject to that of the verb phrase by specifying that the subject is the agent of an event denoted by the verb phrase. This functional category is hierarchically sandwiched between V and T in the extended projection. This proposal decomposes the verb into two syntactic components, the contentful lexical verb *bite* and a functional element notated v (pronounced ‘little v’), that uses an event-semantics representation to add an agent to that verb (Kratzer 1996, Borer 2005, Ramchand 2008 for approaches that execute this idea in diferent ways):

(35)

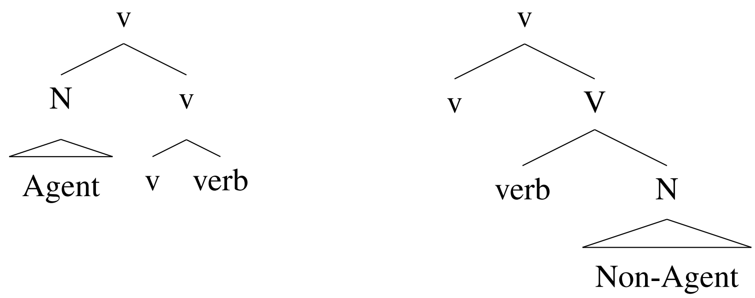


The meaning of v is specified so that *Anson* is interpreted as the agent of *bite Lilly*, while T marks this event as taking place before the speech time. Empirical arguments have been given that in various languages v is overtly expressed by a morpheme (e.g. Mahajan 2012 for Hindi, Coon 2012 for the Mayan language, Chol), although this is an ongoing area of investigation, and much is still unclear. An alternative to this syntactic view is one that takes the verb itself to specify that the participant that is interpreted as the agent is to be syntactically higher, so that the relevant generalization is one about the way that lexical information maps to syntactic information (Grimshaw 1990, Bresnan 2001, Reinhart 2000).

Further work, following on from Dowty (1979), has connected the interpretation of the subject and the object to aspectual properties of the verb (see Folli and Harley 2006 for brief review). The more general outcome of this research is that there is only a small number of basic event types and evidence that these correlate tightly with a small number of basic syntactic structural types (Hale and Keyser 1993).

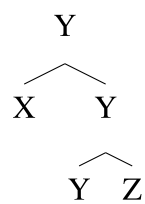
To get a flavour of this, consider the possible syntactic positions for noun phrases that denote participants of the event denoted by a verb (these are usually called the verb’s *arguments*). If the verb has two arguments, as in (35) above, there is no ambiguity about which is which. However, if a verb has just one argument, then there are two possible syntactic positions for that one argument: either where *Lilly* is placed in (35), or where *Anson* is placed. Further, each of these is associated, by hypothesis, with a meaning: the element that combines with the constituent headed by v is associated with an agent meaning, while the sister of the verb is not. We predict two syntactic classes of verb:

(36)



It will be useful to have a terminology for these two different structural positions: we say that the sister of a lexical word or a functional category is the syntactic *complement* of that category, so that the V subtree in the structures above is the complement of v, and the Non-Agent object N is the complement of verb; we then say that the sister of the mother of a lexical word or functional category is its *specifier*, so that the Agent noun in (36) is the specifier of v. More generally, in (37), X is the specifier of Y and Z is the complement of Y, where Y is either a lexical word, like a verb, or a functional category, like v or T. The terms specifier and complement are general ways of talking about structural relationships which cover the traditional notions of subject and object, but go beyond these.

(37)



The two classes in (36) (termed unergative and unaccusative) have been recognized since Perlmutter (1978) and there is a wealth of linguistic evidence that they are structurally, and not just semantically, distinct. The unergative verb *jump* would have the representation on the left, while the unaccusative *fall* would have the structure on the right. In English, this structural difference is masked for independent reasons (see below), however, in other languages, there are numerous structural distinctions dependent on this difference, including the kind of auxiliary that the verb appears with, the possible positions for the noun phrase, and the behaviour of pronominal elements in clause structure (see the introduction to Alexiadou, Anagnostopoulou, and Evereart 2004 for review of the phenomena). There is also solid processing evidence for the distinction from different experimental paradigms: Bever and Sanz (1997) found reaction time differences correlating with the two classes, which they showed to be structural rather than semantic; Friedmann, Taranto, Shapiro, and Swinney (2008) in an online cross modal lexical priming experiment showed priming effects attributable to the structural difference in the position of the verb's argument; Lee and Thompson (2011) showed fixation differences in eyetracking also sensitive to this.

4.4 Syntactic Dependencies

We have seen that expressions of human language have a constituent structure. In addition various dependencies hold between constituents. These dependencies are crucial for determining various aspects of both form and meaning.

4.4.1 Form and Position Related Dependencies

We have already encountered a form–related dependency in section 2: agreement. Agreement is a systematic covariation in the linguistic forms of two or more elements. However, care must be taken as to what we mean by 'form' here. In English examples like (38), it is not the formal plural marking on *girls* (i.e. the *-s*) that covaries with the form of the verb, but rather a more abstract property of this word, which is interpreted as semantic plurality. (38-c) shows that the form of the word is not relevant (as the word 'sheep' does not change its form to mark whether it is singular or plural). It is also not obvious that it is the meaning that is conditioning the agreement, unless we take electronic scales to be plural (but see Dowty and Jacobson 1988 and Pollard and Sag 1994). In other languages, as we will see directly, appeal to meaning is not sufficient to explain agreement.

(38) a. The girls are singing
b. *The girl are singing
c. The sheep is/are bleating
d. The electronic scales are broken

A property of a word or other linguistic expression is called a *feature*: we say that the number feature on the noun (marking its semantics as singular or plural) covaries with a number feature on the verb, which is responsible for the form of the verb.

As well as simple subject verb agreement, languages also display agreement of the verb with objects. For example, the finite verb in the Nakh-Daghestanian language Tsez agrees, in the usual case, with the noun class of its object[3] (Polinsky and Potsdam 2001, example 4; I show the full example here as the actual forms are helpful):

(39) eniy-ā ziya b-išer-si
mother-Erg cow.III III-feed-past.evidential

[3]Noun classes in Tsez are only roughly semantically determined, with class I being male humans, class II being female humans, plus a number of inanimate nouns, class III including all animals and a large set of inanimates, and class IV being inanimate nouns.

“The mother fed the cow.”

Here the verb has a prefix *b*, glossed as III, that covaries with the class of the object noun *ziya*, ‘cow’, glossed here as ‘cow.III’. Compare this with (40) (Maria Polinsky, pers. comm.), where the class changes from III to II with a concommitant change of the prefix to *y*, and (41), where the object number changes from singular to plural and the prefix changes to *r*:

(40) eniy-ā kid y-išer-si
mother-Erg girl.II II-feed-past.evidential
“The mother fed the girl.”

(41) eniy-ā ziya-bi r-išer-si
mother-Erg cow.III-pl III.pl-feed-past.evidential
“The mother fed the cows.”

Some languages allow both subject and object agreement, and in others, agreement can appear for three arguments of a verb (e.g. Basque).

In the cases we have seen so far, agreement is (structurally) local: either a head agrees with a specifier, as in English, or with a complemet, as in Tsez. However, Tsez shows us that this does not have to be the case. In a sentence with an embedded clause, the higher verb can agree with the object of the lower verb, giving the following kind of pattern (Polinsky and Potsdam 2001, example 1b):

(42) mother [boy bread.III ate] III.knows

Here the verb *know* agrees not with its sister constituent (the sentence *boy bread.III ate*) but rather with the object of the verb inside that constituent. This kind of long-distance agreement, although quite rare in languages of the world, is also seen to a certain extent in English examples like:

(43) There **are** likely to be **some girls from Paris** at the party.

Languages also display agreement inside noun phrases, as well as between noun phrases and verbs. For example, in French, articles (words like *the* and *a*) and adjectives systematically covary with the gender (and number) of the noun. French has two words for ‘day’, one masculine in gender and one feminine, and articles and adjectives covary with these:

(44) a. un beau jour
a beautiful day
b. une belle journée
a beautiful day

Another important set of syntactic dependencies involving form are case-dependencies. Take the following sentences in English:

(45) a. He saw her
b. *Him saw her
c. *She saw he
d. She saw him

Here the form of the pronoun is apparently dependent on its syntactic position. The traditional notion of 'case' captures this: nouns and pronouns come in different shapes which are dependent on a case feature they bear, and this case feature is dependent on the syntax. In the examples above, the forms *he*, *she* are *nominative* while the forms *her*, *him* are *accusative*. It turns out that it is not possible to associate case features with meanings directly (as is shown by comparing *She bit me* vs *She was bitten*, where the same case (nominative) appears with different semantic roles (Agent and Non-Agent respectively). Case is connected to structure not to semantics.

However, it is not sufficient to simply associate particular positions in (say) a syntactic tree with certain case forms. In Icelandic, for example, for most verbs, the subject appears in the nominative case (Zaenen, Maling, and Thráinsson 1985). However, for some verbs the subject is non-nominative, and when this happens, the nominative is systematically assigned to the object:

(46) a. SubjectNounPhrase.Nom Verb ObjectNounPhrase.Acc
b. SubjectNounPhrase.Dat Verb ObjectNounPhrase.Nom

This pattern shows that nominative is not associated with the subject position directly. A proposal that has been successful in descriptively caturing some case dependencies is that nominative case is associated with T, and that T and some noun phrase enter into a syntactic dependency, the formal outcome of which is that that noun phrase comes to bear a nominative feature, hence its nominative form. We could schematize this as follows, where [case:nom] abbreviates the idea that the case property is specified as having the value nominative:

(47) T[case:nom] ... NP[case:] $\rightarrow$ T[case:nom] ... NP[case:nom]

There is of course a great deal more to case dependencies than nominative. The astute reader will notice that in the Tsez examples above the subject is marked Erg, which abbreviates ergative case, a case form that, roughly, picks out just agents of two-argument verbs; there is no consensus analysis of ergatives and this is an area of ongoing investigation both empirical and theoretical (Dixon 1994, Woolford 1997, Aldridge 2008, Legate 2012). We have also not discussed accusative case on objects, which has been analysed as having its source not in T but in v (Chomsky 1995, Legate 2008).

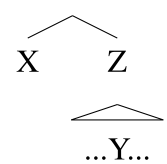
Case and Agreement seem to be related phenomena: they both involve a syntactic dependency set up between a functional category and a noun phrase; they are both dependent on structural not linear relationships; they both involve an abstract feature which can be characterised as being present on both parts of the dependency. In fact, if we look at subject verb agreement in English for number, we can add it to our schema above as follows, where [num:pl] abbreviates a specification that the number agreement property has the value plural:

(48) T[case:nom, num:]...NP[case: , num:pl] $\rightarrow$ T[case:nom, num:pl]...NP[case:nom, num:pl]

The idea would be that the noun phrase is the 'source' of the number feature which T agrees with, while T is the source of the nominative feature which the noun phrase acquires. These features then appear morphologically as agreement and case (see Adger and Svenonius 2011 for syntactic features).

A hypothesis about the structural relationship that is relevant in case and agreement dependencies is that it is the same in both cases, and can be characterised by a relation defined over constituent structure. The standard name for this structural relation is c(onstituent)-command and we can characterize it asa kind of extension of the sister relation: a syntactic element will c-commend its sister and everything inside its sister as follows:

(49)



Here X c-commands its sister Z and everything inside Z, hence Y. Y c-commands its sister and everything inside that sister, hence it does not c-command X. If we take the ellipsis ... in (48) to be c-command, then we would allow T to agree with and assign case to any noun phrase it c-commands. There is a good consensus in the field that c-command, or something very like it, is relevant for understanding how agreement and case form–dependencies are regulated in human languages. Obviously, if this is true, it raises interesting questions about the nature of the sentence processor, and indeed the aetiology of this putative constraint. One influential proposal is that c-command should be replaced with a search-based system, where a feature on X would search through the constituent Z until it reaches a matching feature, at which point the relevant features would be valued (Chomsky 2000, or Adger 2003 for a textbook treatment).

In addition to these dependencies of form, we also see dependencies of position. One such is subcategorization requirements (often termed complement selection). For example, a verb like *hit* requires an object: *Lilly hit the mouse* vs **Lilly hit*. Moreover the object has to be a certain syntactic category (in this case a noun phrase: cf. **Lilly hit that Anson sneezed*, or compare *I depend on you/*I depend you*). Subcategorization dependencies are known to be psychologically very salient, to bear a close but indirect relationship to verb meanings (e.g. Fisher, Gleitman, and Gleitman 1991), but to be not entirely reducable to meaning (Grimshaw 1979). To see this consider:

(50) a. I asked the time
b. I asked what the time was
c. *I inquired the time
d. I inquired what the time was

Here the verbs 'ask' and 'inquire' both combine with a complement whose meaning is a question, giving (50-b) and (50-d). However, only 'ask' can combine with a noun phrase object which has a question meaning, so (49-c) is semantically reasonable but lower in acceptability than the other examples. This pattern suggests that there is a role for purely syntactic information that specifies the kind of object a verb requires (noun phrase or sentence), presumably learned as a feature of the verb in addition to the verb's meaning. Subcategorization requires sisterhood and cannot be non-local, unlike agreement or case dependencies. There are no cases where, for example, a verb specifies the category of some constituent located deeply inside its complement; compare this with agreement, where we saw the Tsez verb agreeing with an object noun phrase inside the verb's sentence complement.

(51) a. Local: subcategorization dependencies hold under sisterhood
b. Non-Local: agreement and case dependencies hold under c-command

Recall in our discussion of unaccusative and unergative verb classes that there was good linguistic and processing evidence for two structural positions: unaccusative verbs' single argument is the sister of the verb, while unergatives' is the specifier of v. This is a subcategorizational difference in verb class: unaccusatives subcategorize an object, unergatives do not. However, in English there is no difference in order, although one might expect the object in the unaccusative to occur after the verb, which is the general order for English:

(52) a. The cat jumped.
b. The cat fell.

The reason for this is another selectional dependency, but this time one that is purely syntactic: in English the specifier of T must be filled by some constituent (the (historically opaque) name for this generalization is EPP). This constraint of English grammar is motivated by a simple distributional fact: we don't find sentences which lack a structural subject in English:

(53) a. (I know that) a cat is in the lavender

b. (I know that) there is a cat in the lavender

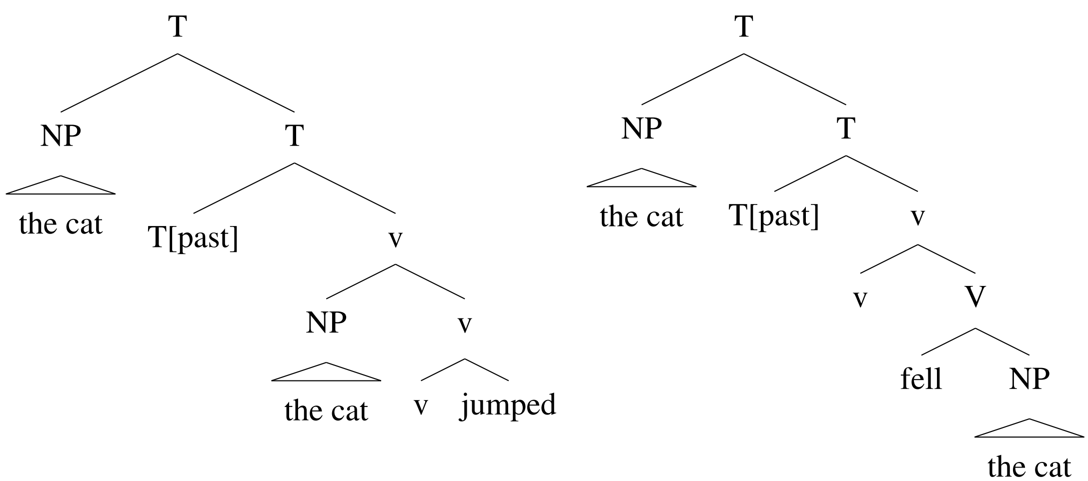
c. *(I know that) is a cat in the lavender

d. It is 4 o’clock/ *is 4 o’clock

(53-b) requires the ‘dummy’ word *there* to be in the position before the auxiliary verb as we can see in (53-c). Similarly, (53-d) requires the pronoun *it*. The distributional generalization is that the specifier of T must be overtly filled in English. Such a requirement is known not to hold in other languages and whether it has a deeper explanation is a question of some controversy. There are also exceptions (for example, imperatives, like *leave now!*) although these are the result of interfering factors, at least in the well-understood cases.

However, taking EPP to be true for English, we now have an explanation for why the surface orders of unaccusatives and unergatives is the same in English. Even though the two verb phrases (technically v phrases) are distinct in structure, this difference is obliterated, because EPP requires the single noun phrase in each structure to appear in the specifier of T position. There are two selectional requirements: one in the verb phrase where the interpretation of the noun phrase as an agent or not is determined, and the other at the specifier of T, which has to be filled in English. These two selectional requirements can both be met by taking the phrase *the cat* to be in both structural positions simultaneously, giving the following representations:

(54)



As mentioned above, there is processing evidence for the difference in the lower structural positions, even though the subject is not pronounced there. These simple examples then have two types of dependency relation: a dependency between the verb and its argument and a distributional regularity of English (EPP) that requires a local syntactic relation between T and some overt specifier. In English, the higher (that is c-commanding) of the two *the cat*s is pronounced, so that we have the effect of a third dependency obtaining between the structural position associated with EPP and the structural position associated with the noun’s interpretation as an agent or non-agent.

This third kind of dependency is a side effect of the other two constraints holding, and it is often termed ‘movement’ or ‘displacement’. When syntacticians talk about movement, what they generally mean is a phenomenon where a single phrase appears to enter into a number of dependencies at once, and further, is pronounced in a single position. The appropriate theoretical model for this phenomenon is orthogonal to its existence, and ubiquity, in languages of the world.

We have seen two form-dependencies (agreement and case), which might be abstractly connected through a requirement that features of items are matched in a c-command relation. We have also seen two position-dependencies: subcategorization for properties of a complement, and EPP. These various dependencies seem not to supervene on semantic properties, but are fundamentally constraints on the form of the structures that human languages allow.

4.4.2 Meaning Related Dependencies

Connected to subcategorization, which requires a particular syntactic category, is semantic selection. Conceived of abstractly, this is the requirement that a certain verb (we restrict attention to verbs here) occur with a defined number of arguments, each of which bears particular meaning relations to the verb. Subcategorized complements, as we have seen, do not bear an Agent relation to their predicate, while the specifier of v does. However, there are more kinds of semantic relationship. Take, for example *Lilly fears Jasper*, where Lilly is not an Agent of fear but rather experiences that emotion, or *Anson baked David a teacake*, where David benefits from the event, although he is not its agent or the entity that undergoes constitutional change in that event. At a more fine-grained level, verbs also impose a certain coherence requirement on their arguments, which is what leads to the bizarreness of examples like *the amoeba coughed*, *the proposal kicked me*, etc., which require metaphorical or dream contexts for interpretation. These various types of meaning dependencies centre around the verb and appear to be local to the verb, although not as local as syntactic subcategorization. There is growing evidence for distinct cortical activations for subcategorization versus semantic selection, confirming the linguistic distinctions that have been proposed (Ardila 2012, Yang, Khodaparast, Bradley, Fang, Bernstein, and Krawczyk 2013).

Perhaps the more striking meaning-related dependencies which have structural correlates involve what are called ‘bound variable effects’. The following example has two distinct meanings:

(55) Every kitten wanted its food.

The most salient meaning requires that kitten A wants the food that belongs to kitten A, while kitten B wants the food that belongs to kitten B etc. This is the bound variable reading, as it can be represented in (Anglicized) predicate logic as (56), where the variable x is bound by the quantifier *every*:

(56) For every x, where x is a kitten, x wants x's food.

The other meaning is that there is some other individual (say a puppy) whose food we are talking about:

(57) a. The puppy tore into the steak it had stolen. Every kitten wanted its food.
b. There is some y and for every x, where x is a kitten, x wants y's food.

When we change the structural relationship between the phrase *every kitten* and the pronoun *its*, the bound variable reading vanishes (or at least becomes much less accessible)

(58) Its food was in every kitten's bowl.

This phenomenon is highly complex, and structural conditions are possibly only part of what is going on (Barker 2012). The question of how constituent structure interlinks with linear order and with information and discourse structure in identifying the meaning of a pronoun is still wide open. However, the semantic relation we see here is extremely important and reified in some very clear ways elsewhere in syntax.

One important case is reflexives. Consider:

(59) Lilly impresses herself.

The meaning of this requires the two arguments of the verb *impress* to be construed as the same individual. That is, the verb phrase *impresses herself* means something like *x, such that x impresses x*, where we fill in the variable x with the meaning of the noun *Lilly*. We see a similar effect with the expression *each other*:

(60) Lilly and Jasper impressed each other

Here the verb phrase means something like *x and y, such that x impresses y and y impresses x*.

There are clear structural conditions on recovering this meaning from sentences. For example, a reflexive in English requires that the expression which supplies the meaning (its antecedent) is in a certain structural relationship with it. The following cannot mean that Lilly is such that her owner impresses Lilly. I have placed brackets in (61) to mark the constituent structure:

(61) [[[Lilly's] owner] [impresses herself]]

One hypothesis about the structural relationship between the reflexive and its antecedent is that the latter must c-command the former. In the bracketting above, *Lilly* has the word *owner* as its sister, and hence cannot c-command the reflexive *herself*. If the bound variable meaning of the reflexive requires that it's antecedent c-commnd it, this will follow. This hypothesis is attractive, as c-command appears to be relevant for other cases of form-meaning relations, and it extends to a very wide range of data across languages. However, there are cases where it is not sufficient to model the behaviour of reflexives and various extra proposals have been adduced to deal with their full distribution (see Safir 2013 for review).

Another case where we find a bound variable semantics is resumptive pronouns. These are found in many language families (Celtic, Semitic, Kru, Germanic, Polynesian), and appear in structures that look as follows:

(62) Which girl did the cat scratch her?

The meaning of this question is roughly: *for which x, where x is a girl, is it the case that the cat scratched x*. The pronoun and the constituent *which girl* are construed together, giving us a bound variable semantics. McCloskey (2006) provides a survey of the empirical range of resumptive pronoun structures and their theoretical analysis.

The equivalent of (62) in English is of course (63):

(63) Which girl did the cat scratch?

This also has a bound variable reading, but there is no pronoun or other overt element to mark the position of the variable as the object of the verb *scratch*. Instead there is simply the absence of the object, often termed a gap. This same pattern appears in many distinct places in the grammar of English and other languages. (64) shows it also in relative clauses and topicalizations (Chomsky 1977 for a much wider range of cases):

(64) a. The mouse that the cat bit.
b. That mouse, the cat will never catch.

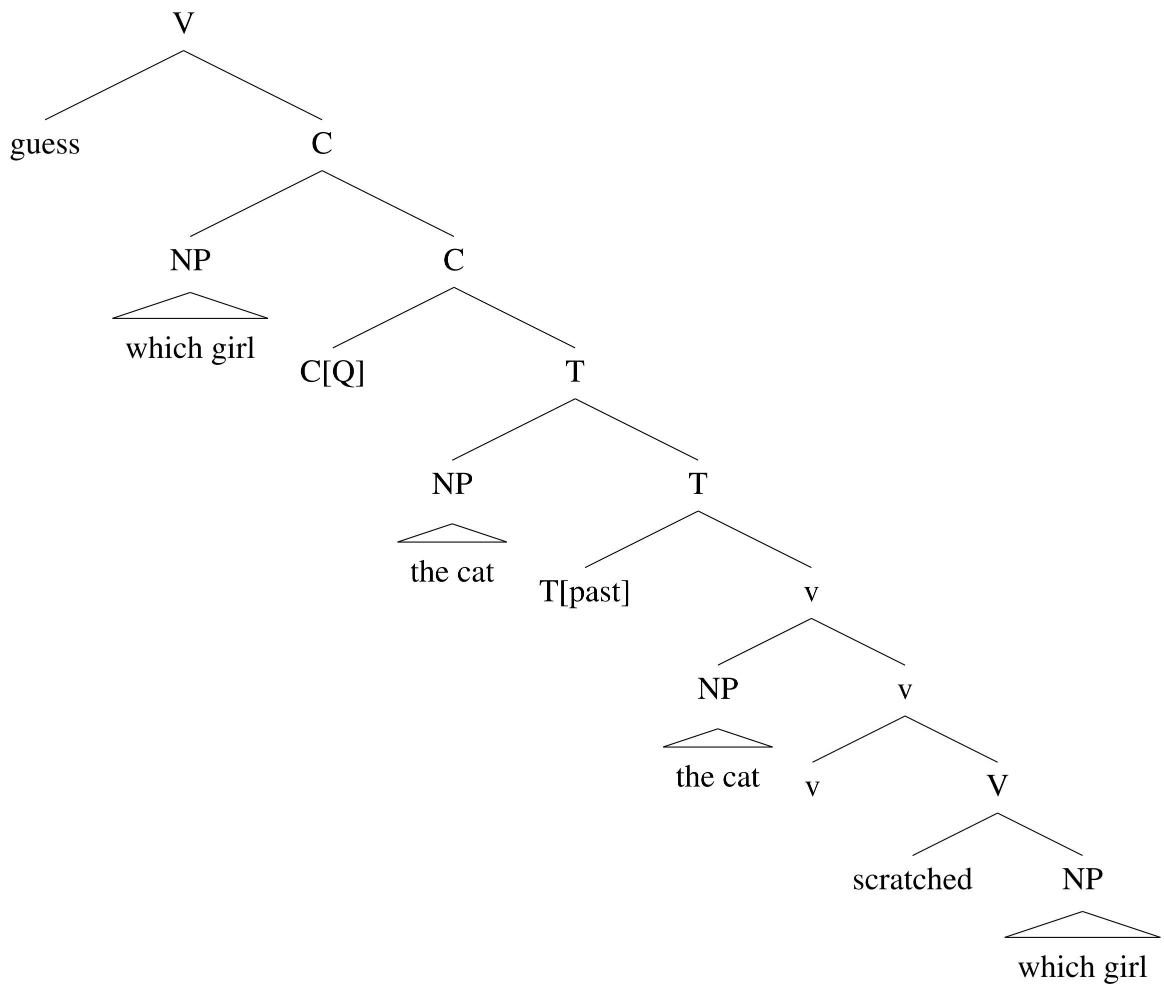
One possible analysis is that there is an unpronounced pronoun in the position of the object, so that English works just like a language with resumptive pronouns. However, both linguistic and experimental evidence has shown that gaps do not behave like resumptives (McCloskey 2006, Suñer 1998 for linguistic and Alexopoulou 2006, Heestand, Xiang, and Polinsky 2011 for experimental evidence).

The construction exemplified in (63) is termed a constituent question, since the constituent at its beginning is the focus of the question. This construction has spawned a vast literature because of the linguistic and processing issues that it raises. It is found in matrix clauses and also in embedded clauses:

(65) Guess which girl the cat scratched!

In many languages, the constituent that is questioned is specially marked; in English the relevant marking often involves words containing *wh*, hence the other name for these questions: *wh*-questions. In English, there is a positional dependency set up between the functional category C, when it has a question interpretation, and a *wh*-constituent, such that the question interpretation arises when the *wh*-constituent is the specifier of C. Just as we saw with EPP in English, this means that there are two distinct positional dependencies in the sentence: the *wh*-constituent is required to be local to the verb *scratch*, in fact in this sentence it must be the complement of *scratch* to get the interpretation correct; at the same time the *wh*-constituent must be in the specifier of C. The relevant structure that models this is therefore:

(66)



This structure encodes the semantic relationships between the verb and its arguments, the syntactic relation between the EPP requirement of T in English, and the semantically interpreted positional dependency between the question version of C (C[Q]) and its specifier. Just as we saw for the EPP, only the c-commanding token of *which girl* is pronounced, but both are interpreted semantically: the c-commanding one is interpreted as a quantifier (roughly, *for which x, where x is a girl*) and the c-commanded token is interpreted as a bound variable (in this case the variable *x* bound by the quantifier).

Not all languages require the specifier of C[Q] to be overtly filled to make a constituent question. Many languages (e.g. Japanese, Chinese) have the *wh* constituent in its position inside the sentence. However, there is still strong evidence for a linguistic and psychological relationship between the *wh*-constituent and C[Q]. Richards (2010) shows that a particular prosodic structure is associated with the span of elements between the *wh*-expression and the C, while Ueno and Kluender (2009) found a reliable neural processing correlate of this relation in Japanese. One analytical option here would be to say that such languages have a representation very similar to (66), but that the lower of the two tokens of the *wh*-constituent is pronounced, rather than the higher. Again, the theoretical and analytical issues are under debate, but the core empirical generalizations are not.

It is clear that bound variable dependencies are non local. Moreover, for the cases surveyed here, c-command, or something very similar to it, appears to be required. As a descriptive generalization, we do not find languages where a *wh*-constituent does not c-command the position where it is interpreted as a variable, suggesting that it may be possible to unify the structural conditions on all syntactic dependencies, although this is some way off.

One final point about bound variable dependencies is important. Although they are restricted by c-command, phenomenologically, at least, they appear to be unbounded:

(67) Which problem did you intuit [that Anson would say [his administrator would persuade [Morag to solve]]]?

Here the gap to be associated with *which problem* is separated from it by three clauses (marked in the example with brackets). Although memory issues quickly intervene, no one has ever established a systematic limit on how many sentence boundaries this *wh* dependency can cross, and the class of dependencies that work this way are often refered to as unbounded dependencies (or long-distance dependencies). However, there is linguistic evidence that the C position after each verb c-commanded by the *wh*-phrase (in this case, *intuit*, *say*, *persuade*, *solve*) is part of the dependency (McCloskey 2002, Chung 1994, Georgopoulos 1985, Adger and Ramchand 2005) and experimental evidence that the human sentence processor posits a gap (interpreted as a bound variable) in these positions (Gibson and Warren 2004). This has been taken as evidence that apparently long-distance dependencies like this are composed of a series of shorter dependencies strung together (as first suggested by Chomsky 1973).

Connected to this issue, even long distance dependencies are subject to certain locality effects (Ross 1967), known as Island Conditions. For example, although a *wh*-constituent may link to its gap over intervening sentences, it may not if the intervening sentence is itself a subject, or is inside a noun phrase. In (68), *that Lilly scratched Anson* is the subject or the verb *surprise*, but it is impossible to make a parallel constituent question where the object of *scratch* is in the specifier of C; in (69) the object *the mouse* contains a modifying sentence (a relative clause) which itself contains a further constituent *Anson*. Again, this constituent cannot be questioned:

(68) a. [That Lilly scratched Anson] surprised David.
b. *Who did that Lilly scratched surprise David

(69) a. Lilly caught [the mouse that Jasper brought Anson]
b. *Who did Lilly catch the mouse that Jasper brought.

There are many Island configurations, and an enormous amount of work has gone into understanding their nature, their cross-linguistic validity, and their explanation (Boeckx 2012, den Dikken and Lahne 2013, Sprouse and Hornstein 2013). Whether they are to be understood as emerging from the structure of the grammar (Chomsky 1973) or from the structure of other cognitive processes (such as the sentence processor, Kluender and Kutas 1993) is a current topic of debate (Sprouse, Wagers, and Phillips 2012), but the empirical phenomenon is robust and clearly important to cognitive science vas a whole.

This section has sketched only a few of the most important kinds of syntactic dependencies found in human language. Others involve obligatory interpretation of the arguments of different predicates as the same entity, as in (70-a), where the understood subject of *catch* is interpreted as identical to the subject of *attempt*; apparent anti-dependencies where a pronoun and a noun phrase cannot be construed as the same (as *she* and *Lilly* in (70-b)); and cases where there is a single *wh* constituent, but multiple gaps, as in (70-c), where *which mouse* is understood as the object of both *sratch* and *ate*. Of course, these are just examples from English, while syntax investigates how or if they appear in other languages and what consequences the variation and uniformity across languages has for understanding the nature of the human syntactic capacity.

(70) a. Lilly attempted to catch the mouse.
b. She didn't understand that Lilly was sleepy.
c. Which mouse did Lilly scratch before she ate?

5 Theoretical models

A lurking suspicion in the minds of some psychologists or cognitive scientists may be that syntactic theory need not exist as part of the explanation of human language because there is nothing for it to be a theory of. It is worth briefly considering what this would entail.

There is a very strong form of this idea that no one endorses: it would say that even adults have no cognitive store of information relevant to the structures that link sound and meaning. No one adopts this view, because, in the absence of syntactic information, we lose any explanation for even the most basic properties of language sketched above.

A weaker version of the claim would be that human neonates have no cognitive store of information relevant to human language, and further no systems of learning specialised to language. The cognitive store of information about syntax that a developing human comes to have and use is built from other systems that are not language specific (cf. Tomasello 2003 and other variants of usage-based grammatical theories), so our theoretical efforts should be directed towards finding out how these other systems work.

Any learning mechanism M will map from the domain of possible linguistic experience (D) to the domain of possible generalizations about the data (G), where G allows the learner to interpret data it has not previously encountered. This characterization holds, irrespective of whether this mechanism is syntax-specific, usage-based or whatever:

(71) M: D $\rightarrow$ G

Classical work in learning theory (Gold 1967, Horning 1969, Niyogi 1998) concludes that figuring out what G is from D is an extremely difficult problem. One solution to the problem is that the form of the generalizations that G encodes is constrained; that is, there are characterizable limits to the class of grammars—this is the solution proposed by theoretical syntax. Could we deny this, and propose instead that the characterization of M itself is sufficient to limit the possible grammars, where the components of M are not specialized to syntax? In such an approach, the explanatory contribution has to be in specifying M, not G. However, M has to be specified in such a way that we can account, for example, for the insensitivity of language to local linear context: that is, somehow M has to know not to apply a whole slew of different kinds of learning algorithm to language. But this is of course equivalent to saying that M partitions the hypothesis space in a particular fashion so that linguistic data is acquired differently from other data, hence there are systems of learning specialised to language. The further question is then whether those systems are also at play elsewhere in cognition, a question that is wide open. The hunch within syntax is that our increasing understanding of syntactic structure will lead to interesting models for other aspects of cognition, in the same way that our understanding of cognition in general can lead to a deepening and potential simplification of our models of syntax.

There is also excellent empirical evidence that there are limits on the class of grammars. Although the discipline of theoretical syntax can look both factional and fractious, with little consensus, that is in fact not at all the case. A number of apparently quite different syntactic theories, when appropriately formalized, all turn out to have some straightforward mathematical similarities, as first suggested by Joshi (1985) and recently emphasized by Stabler (2013). Such theories can generate a class of linguistic structures that is larger than those of context free grammars, but smaller than those of context-sensitive grammars, which in turn are smaller than recursively enumerable grammars (Chomsky 1956). The current consensus, then, is that the power of models such as the transformational model proposed in Chomsky (1965), or the unification based model proposed in Pollard and Sag (1994) is not required to analyse human language, while the context free model of Gazdar, Klein, Pullum, and Sag (1985) is too limited. The 'Goldilocks' class include Tree-Adjoining Grammars (Joshi 1985), certain versions of Combinatory Categorial Grammars (Steedman 1996) and Minimalist Grammars (Stabler 1998), as well as extended versions of Context Free Grammars such as Multiple Context Free Grammars (Pollard 1984). Intuitively, these approaches are similar in how they account for a particular kind of linguistic dependency (so called cross-serial dependencies), as well as in being efficiently parsable, and in having a certain similarity in the class of derivations they allow. It is unclear how this result can be characterized, or would have ever been discovered, in an approach that a priori places no limits on the class of grammars in general. In fact, syntactic research has shown that the class of actual human languages is yet more restricted than these limits. We have discussed what kinds of further restrictions apply in section 4.

References

Abels, Klaus and Neeleman, Ad. 2012. Linear asymmetries and the LCA. *Syntax* 15:25–74.

Adger, David. 2003. *Core Syntax*. Oxford: Oxford University Press.

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

Adger, David, Harbour, Daniel, and Watkins, Laurel. 2009. *Mirrors and Microparameters: Phrase Structure beyond Free Word Order*. Cambridge: Cambridge University Press.

Adger, David and Ramchand, Gillian. 2005. Move and Merge: *Wh*-dependencies revisited. *Linguistic Inquiry* 36:161–194.

Adger, David and Svenonius, Peter. 2011. Features in minimalist syntax. In Cedric Boeckx, ed., *Oxford Handbook of Linguistic Minimalism*, 27–51, Oxford, UK: OUP.

Aldridge, Edith. 2008. Generative approaches to ergativity. *Language and Linguistics Compass* 2:966–995.

Alexiadou, Artemis, Anagnostopoulou, Elena, and Evereart, Martin, eds. 2004. *The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface*. Oxford: Oxford University Press.

Alexopoulou, Theodora. 2006. Resumption in relative clauses. *Natural Language and Linguistic Theory* 24:57–111.

Anderson, Stephen and Chung, Sandra. 1977. On grammatical relations and clause structure in verb-initial languages. In Peter Cole and Jerry Sadock, eds., *Grammatical relations*, Syntax and semantics 8, 1–25, New York: Academic Press.

Ardila, Alfredo. 2012. Interaction between lexical and grammatical language systems in the brain. *Physics of Life Reviews* 9:198 – 214.

Barker, Chris. 2012. Quantificational binding does not require c-command. *Linguistic Inquiry* 43:614–633.

Berwick, Robert C., Friederici, Angela D., Chomsky, Noam, and Bolhuis, Johan J. 2013. Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* 17:89 – 98.

Bever, Thomas G. and Sanz, Montserrat. 1997. Empty categories access their antecedents during comprehension: Unaccusatives in Spanish. *Linguistic Inquiry* 28:pp. 69–91.

Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18:355–387.

Bock, Kathryn, Eberhard, Kathleen M., Cutting, J.Cooper, Meyer, Antje S., and Schriefers, Herbert. 2001. Some attractions of verb agreement. *Cognitive Psychology* 43:83 – 128.

Boeckx, Cedric. 2012. *Syntactic Islands*. Cambridge: Cambridge University Press.

Borer, Hagit. 2005. *The Normal Course of Events: Structuring Sense*. Oxford University Press.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Brody, Michael. 2000. Mirror theory: syntactic representation in perfect syntax. *Linguistic Inquiry* 31:29–56.

Carnie, Andrew. 2010. *Constituent Structure*. Oxford, UK: Oxford University Press.

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

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

Chomsky, Noam. 1968. *Language and Mind*. New York: Harcourt, Brace and World.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson and Paul Kiparsky, eds., *A Festschrift for Morris Halle*, New York: Holt, Rinehart, and Winston.

Chomsky, Noam. 1976. Conditions on the rules of grammar. *Linguistic Analysis* 2:303–351, reprinted in **?**).

Chomsky, Noam. 1977. On *wh*-movement. In P. Culicover, T. Wasow, and A. Akmajian, eds., *Formal Syntax*, 71–132, New York: Academic Press.

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

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

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In R. Martin, D. Michaels, and Juan Uriagereka, eds., *Step by Step: Essays on Minimalist syntax in Honour of Howard Lasnik*, 89–115, Cambridge, MA: MIT Press.

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

Chung, Sandra. 1994. Wh-agreement and "referentiality" in Chamorro. *Linguistic Inquiry* 25:1–44.

Chung, Sandra. 2005. Properties of VOS languages. In Martin Everaert, Henk van Riemsdijk, Rob Goedemans, and Bart Hollebrandse, eds., *The Blackwell Companion to Syntax*, 685–720, Blackwell.

Cinque, Guglielmo. 1999. *The Syntax of Adverbs*. Oxford: Oxford University Press.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36:315–332.

Cinque, Guglielmo and Kayne, Richard, eds. 2008. *The Oxford Handbook of Comparative Syntax*. Cambridge: Oxford University Press.

Cinque, Guglielmo and Rizzi, Luigi. 2010. The cartography of syntactic structures. In Bernd Heine and Heiko Narrog, eds., *The Oxford Handbook of Linguistic Analysis*, 51–66, Oxford University Press.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cohen, Laurent and Mehler, Jacques. 1996. Click monitoring revisited: An on-line study of sentence comprehension. *Memory and Cognition* 24:94–102.

Collins, Chris. 2002. Eliminating labels. In Samuel David Epstein and T. Daniel Seeley, eds., *Derivation and Explanation in the Minimalist Program*, 42–64, Blackwell Publishing.

Coon, Jessica. 2012. Split ergativity and transitivity in chol. *Lingua* 122:241–256.

Culicover, Peter and Jackendoff, Ray. 2010. Quantitative methods alone are not enough: Response to Gibson and Fedorenko. *Trends in Cognitive Science* 14:234–235.

den Dikken, Marcel, ed. 2013. *The Cambridge Handbook of Generative Syntax*. Cambridge: Cambridge University Press.

den Dikken, Marcel and Lahne, Antje. 2013. The locality of syntactic dependencies. In Marcel den Dikken, ed., *The Cambridge Handbook of Generative Syntax*, 655–697, Cambridge University Press.

Diaz, Michele T. and McCarthy, Gregory. 2009. A comparison of brain activity evoked by single content and function words: An fMRI investigation of implicit word processing. *Brain Research* 1282:38 – 49.

Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Dowty, David and Jacobson, Pauline. 1988. Agreement as a semantic phenomenon. In John Powers and Franscisca de Jong, eds., *Proceedings of the Fifth ESCOL*, 95–101, Ithaca: Cornell Linguistics Club.

Dowty, David R. 1979. *Word meaning and Montague grammar*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Emonds, Joe. 1978. The verbal complex $\text{V}' - \text{V}$ in French. *Linguistic Inquiry* 9:151–175.

Erlewine, Michael Yoshitaka and Kotek, Hadas. 2013. A streamlined approach to online linguistic surveys. http://ling.auf.net/lingbuzz/001802.

Everaert, Martin, Riemsdijk, Henk Van, Goedemans, Rob, and Hollebrandse, Bart, eds. 2005. *The Blackwell Companion to Syntax, Volumes I–V*. Oxford and New York: Wiley-Blackwell.

Ferguson, Charles. 1996. Variation and drift: Loss of agreement in Germanic. In Gregory Guy, Crawford Feagin, Deborah Schiffrin, and John Baugh, eds., *Towards a Social Science of Language: Papers in honor of William Labov. Volume 1, Variation and Change in Language and Society*, 173–198, Amsterdam: John Benjamins.

Fillenbaum, Samuel. 1971. Psycholinguistics. *Annual Review of Psychology* 11:251–308.

Fisher, Cynthia, Gleitman, Henry, and Gleitman, Lila R. 1991. On the semantic content of subcategorization frames. *Cognitive Psychology* 23:331 – 392.

Fodor, Jerry, Bever, Tom, and Garrett, Merril. 1974. *The Psychology of Language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Folli, Raffaella and Harley, Heidi. 2006. What language says about the psychology of events. *Trends in cognitive sciences* 10:91–92.

Frederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. 2006. The brain differentiates human and non-human grammars : functional localization and structural connectivity. *Proceedings of the National Academy of Sciences* 103:2458–2463.

Friederici, Angela D. 2011. The brain basis of language processing: from structure to function. *Physiological reviews* 91:1357–1392.

Friedmann, Naama, Taranto, Gina, Shapiro, Lewis P., and Swinney, David. 2008. The leaf fell (the leaf): The online processing of unaccusatives. *Linguistic Inquiry* 39:355–377.

Fromkin, Victoria A. 1971. The nonanomalous nature of anomalous utterances. *Language* 47:27–52.

Gazdar, Gerald, Klein, Ewan, Pullum, Geoff, and Sag, Ivan. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwells.

Georgopoulos, Carol. 1985. Variables in Palauan syntax. *Natural Language and Linguistic Theory* 3:59–94.

Gervain, Judit and Mehler, Jacques. 2011. Speech perception and language acquisition in the first year of life. *Annual Review of Psychology* 61:191–218.

Gibson, Edward and Fedorenko, Evelina. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 16:233–234.

Gibson, Edward., Piantadosi, Steve, and Fedorenko, Kristina. 2011. Using Mechanichal Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5:509–524.

Gibson, Edward and Warren, Tessa. 2004. Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7:55–78.

Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10:447–474.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg, ed., *Universals of Language*, 73–113, Cambridge: MIT Press.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10:279–326.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: MIT Press.

Grimshaw, Jane. 1991. Extended projections, ms, Brandeis University.

Haegeman, Liliane. 1991. *Introduction to Government and Binding Theory*. Oxford: Blackwell.

Hale, Ken and Keyser, Samuel J. 1993. On argument structure and the lexical expression of syntactic relations. In Ken Hale and Samuel J. Keyser, eds., *The View from Building 20*, 53–109, Cambridge, Massachusetts: MIT Press.

Hale, Kenneth. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 1:5–47.

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

Heestand, Dustin, Xiang, Ming, and Polinsky, Maria. 2011. Resumption still does not rescue islands. *Linguistic Inquiry* 42:138–152.

Heim, Irene and Kratzer, Angelika. 1998. *Semantics in Generative Grammar*. Oxford, UK: Blackwells.

Henry, Alison. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford: Oxford University Press.

Hochamnn, Jean-Rémy. 2013. Word frequency, function words and the second gavagai problem. *Cognition* 128:13–25.

Holmberg, Anders. 2013. Verb second. In Artemis Alexiadou and Tibor Kiss, eds., *Syntax–A Handbook of International Research*, Mouton de Gruyter.

Horning, James J. 1969. A study of grammatical inference. Ph.D. thesis, Stanford University.

Hudson, Richard. 2010. *An Inroduction to Word Grammar*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Joshi, Avarind. 1985. Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In David Dowty, Lauri Karttunen, and Arnold Zwicky, eds., *Natural Language Parsing*, 206–250, Cambridge University Press.

Kayne, Richard. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Keenan, Edward and Stabler, Edward. 2004. *Bare Grammar: Lectures on Linguistic Invariants*. Stanford: CSLI.

Kim, Karl HS, Relkin, Norman R, Lee, Kyoung-Min, and Hirsch, Joy. 1997. Distinct cortical areas associated with native and second languages. *Nature* 388:171–174.

Kluender, Robert and Kutas, Marta. 1993. Subjacency as a processing phenomenon. *Language and cognitive processes* 8:573–633.

Koopman, Hilda. 1984. *The syntax of verbs. From verb movement rules in the Kru languages to universal grammar*. Studies in generative grammar 15, Dordrecht: Foris.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck and Laurie Zaring, eds., *Phrase Structure and the Lexicon*, 109–138, Dordrecht: Kluwer.

Lee, Jiyeon and Thompson, Cynthia K. 2011. Real-time production of unergative and unaccusative sentences in normal and agrammatic speakers: An eyetracking study. *Aphasiology* 25:813–825.

Legate, Julie. 2002. Warlpiri: Theoretical implications. Ph.D. thesis, MIT.

Legate, Julie Anne. 2008. Morphological and abstract case. *Linguistic Inquiry* 39:55–101.

Legate, Julie Anne. 2012. Types of ergativity. *Lingua* 122:181 – 191, ¡ce:title¿Accounting for Ergativity¡/ce:title¿.

Levelt, Willem J. M. 1970. Hierarchical chunking in sentence processing. *Perception and Psychophysics* 8:99–103.

Lyons, John. 1968. *Theoretical Linguistics*. Cambridge: Cambridge University Press.

Mahajan, Anoop. 2012. Ergatives, antipassives and the overt light v in hindi. *Lingua* 122:204–214.

McCloskey, James. 1983. A VP in a VSO language? In Gerald Gazdar, Ewan Klein, and Geoff Pullum, eds., *Order Concord and Constituency*, 9–55, Dordrecht: Foris.

McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. In Samuel D. Epstein and T. Daniel Seeley, eds., *Derivation and Explanation in the Minimalist Program*, 184–226, London: Blackwell.

McCloskey, James. 2006. Resumption. In *The Blackwells Companion to Syntax*, 94–117, Oxford: Blackwells.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., and Weiller, C. 2003. Broca’s area and the language instinct. *Nature Neuroscience* 6:774–781.

Muysken, Pieter. 2009. Functional categories: An evolutionary perspective. In Rudolf Botha and Henriette de Swart, eds., *Language Evolution: the view from resticted linguistic systems*, 179–202, Utrecht: LOT.

Muysken, Pieter. 2011. *Functional Categories*. Cambridge: Cambridge University Press.

Myers, James. 2009. Syntactic judgment experiments. *Language and Linguistics Compass* 3:406–423.

Newport, Elissa L. 1990. Maturational constraints on language learning. *Cognitive Science* 14:11–28.

Niyogi, Partha. 1998. *The informational complexity of learning: perspectives on neural networks and generative grammar*. Boston: Kluwer Academic Publishers.

Penke, Martina. 2011. Syntax and language disorders. In Tibor Kiss and Artemis Alexiadou, eds., *Syntax–An International Handbook of Contemporary Research*, Syntax, Walter De Gruyter.

Perlmutter, David. 1978. Impersonal passives and the unaccusative hypothesis. In *Proceedings of the Fifth Meeting of the Berkley Linguistics Society*, UC Berkley.

Pesetsky, David. 1995. *Zero Syntax: Experiencers and Cascades*. Cambridge, MA: MIT Press.

Phillips, Colin. 1996. Order and structure. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Pickering, Martin J. and Ferreira, Victor S. 2008. Structural priming: A critical review. *Psychological Bulletin* 134:427–459.

Polinsky, Maria and Potsdam, Eric. 2001. Long distance agreement and topic in Tsez. *Natural Language and Linguistic Theory* 19.

Pollard, Carl. 1984. Generalized phrase structure grammar. Ph.D. thesis, Stanford University.

Pollard, Carl and Sag, Ivan. 1994. *Head Driven Phrase Structure Grammar*. Stanford: CSLI.

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press.

Reinhart, T. 2000. *The Theta System: Syntactic Realization of Verbal Concepts*. Utrecht: UiL OTS Working Papers.

Richards, Norvin. 2010. *Uttering Trees*. Cambdridge, MA: MIT Press.

Rizzi, Luigi. 2013. Introduction: Core computational principles in natural language syntax. *Lingua* 130:1 – 13.

Roberts, Ian. 2005. *Principles and Parameters in a VSO language*. Oxford and New York: Oxford University Press.

Ross, John. 1967. Constraints on variables in syntax. Ph.D. thesis, Massachusetts Institute of Technology.

Ross, John. 1969. Auxiliaries as main verbs. In *Studies in Philosophical Linguistics, Series One*, Evanston, Illinois: Great Expectations Press.

Safir, Ken. 2013. Syntax, binding and patterns of anaphora. In Marcel den Dikken, ed., *The Cambridge Handbook of Generative Syntax*, 515–576, Cambridge University Press.

Schütze, Carson and Sprouse, Jon. 2013. Acceptability judgments. In Robert J. Podesva and Devyani Sharma, eds., *Research Methods in Linguistics.*, Cambridge:: Cambridge University Press.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle, and Alan Yu, eds., *The Handbook of Phonological Theory*, Oxford: Blackwell Publishers.

Sproat, Richard. 1985. Welsh syntax and VSO structure. *Natural Language and Linguistic Theory* 3:173–216.

Sprouse, Jon and Almeida, Diogo. 2013. The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Provesses* 28:222–228.

Sprouse, Jon and Hornstein, Norbert, eds. 2013. *Experimental Syntax and Island Effects*. Cambridge: Cambridge University Press.

Sprouse, Jon, Wagers, Matt, and Phillips, Colin. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88:82–123.

Stabler, Edward. 1998. Acquiring languages with movement. *Syntax* 2:72–97.

Stabler, Edward P. 2013. Two models of minimalist, incremental syntactic analysis. *Topics in Cognitive Science* 5:611–633.

Steedman, Mark. 1990. Gapping as constituent coordination. *Linguistics and Philosophy* 13:207–264.

Steedman, Mark. 1996. *Surface structure and interpretation*. Cambridge, Massachusetts: MIT Press.

Steedman, Mark and Baldridge, Jason. 2011. Combinatory categorial grammar. In *Non-Transformational Syntax*, 181–224, Oxford, UK: Blackwell Publishers.

Stowell, Tim. 1981. Origins of phrase structure. Ph.D. thesis, MIT, Cambridge, MA.

Suñer, Margarita. 1998. Resumptive restrictive relatives: A crosslinguistic perspective. *Language* 74:335–364.

Tomasello, Michael. 2003. *Constructing a Language: a Usage-Based Theory of Language Acquisition*. Cambdridge, MA: Harvard University Press.

Tsujimura, Natsuko. 2013. *An Introduction to Japanese Linguistics*. Oxford: Blackwell.

Ueno, Mieko and Kluender, Robert. 2009. On the processing of japanese wh-questions: An {ERP} study. *Brain Research* 1290:63 – 90.

Wells, R. 1947. Immediate constituent analysis. *Language* 23:81–117.

Woolford, Ellen. 1997. Four-way case systems: Ergative, nominative, objective and accusative. *Natural Language and Linguistic Theory* 15:181–227.

Yang, Fan–pei Gloria, Khodaparast, Navid, Bradley, Kailyn, Fang, Min-Chieh, Bernstein, Ari, and Krawczyk, Daniel C. 2013. The influence of semantic property and grammatical class on semantic selection. *Brain and Language* 124:194 – 203.

Zaenen, Annie, Maling, Joan, and Thráinsson, Hóskuldur. 1985. Case and grammatical functions: The icelandic passive. *Natural Language and Linguistic Theory* 3:441–483, cited By (since 1996)48.