1 Syntactic Data, Patterns, and Structure

1.1 Introduction: The Problem, and a Possible Solution

1.1.1 Natural Language Data

Linguists appear to be in an enviable position among scientific disciplines. The lifeblood of science is data, and unlike, say, glaciologists, who can only collect primary material for their research in remote and generally rather inhospitable parts of the planet, or particle physicists, who require access to massive, extremely expensive, and sometimes erratic machines – with demand for access to such machines far exceeding supply – linguists are literally surrounded by the kind of data that make up the target of their investigations. It's true that field linguists need informants at a considerable distance from where they themselves live, and experimental linguists often need laboratories with elaborate and sophisticated equipment. But for syntacticians – linguists who investigate the structure of sentences, a large fraction of whom (possibly a majority) study sentences in their own respective languages – matters are as convenient as they could possibly be. Syntacticians have intuitive access to all of the sentences made available by their own knowledge of their language, as well as the speech (and reactions) of their fellows in constant use around them, and ready-made corpora in the form of written materials and electronic records, many of which are available for searches based on word sequences (Google, for example, is a valuable source of data for both syntacticians and morphologists). Learning how to take advantage of this vast pool of readily available data is a major component of syntacticians' training.

In a sense, of course, the true data of syntax are not strings of words themselves, but *judgments* about the status of those strings of words. The syntactician's primary responsibility is to give an account of how it is that certain strings of words have the status of sentences, while other do not, and still others have a kind of shadowy intermediate status – not bad enough to be outright rubbish, but not good enough to pass completely unnoticed in conversation as utterly and tediously normal. For example, consider the status of the three word strings in (1):

(1) a. I asked Robin to leave the room.
b. I requested Robin to leave the room.
c. I inquired (of) Robin to leave the room.

Many speakers appear to rank (1)a as completely unexceptionable, (1)b as 'off' in some noticeable way but not totally unacceptable, and (1)c to be altogether wrong, something that a native speaker of English would simply never say. A complete account of what gives a particular string of words the status of a sentence should, ideally, provide an explanation for the peculiar nature of the deviance of (1)b and the altogether ill-formed status of (1)c. In constructing such an explanation, the linguist will bear very much in mind that *ask* and *inquire* can substitute for each other quite acceptably in the sentences in (2).

(2) a. I $\left\{\begin{array}{c}\text{asked}\\ \text{inquired}\end{array}\right\}$ about Robin's health.

b. I $\left\{\begin{array}{c}\text{asked}\\ \text{inquired}\end{array}\right\}$ whether Robin would be home later.

What is going on in (1) to produce the striking disparity between the first and third examples?

One obvious approach we might start with is that on one of its senses, *ask* points to a relationship between (i) a speaker, (ii) a potential actor, and (iii) an action, of the same general type as *command* or *ordered*, but much weaker. In all of these cases, the speaker has expressed a solicitation for Robin to carry out a certain action, one whose precise nature is made clear by the sequence of words *to leave the room*. That preference may have relatively little force to back it up, or, in the case of *command*, quite a bit of force. *Inquire*, on the other hand, does not seem to correspond to this sense of *ask*, but only to a solicitation of information. So one might assume that the contrast between the first and third examples in (1) – which itself is conspicuously at variance with the verbs' overlapping distribution in (2) – reflects the fact that only verbs which correspond to some solicitation of action can appear in a context, such as (1), where action, rather than information, is sought.

This kind of account is intuitively satisfying, because it seems to correspond to common sense: using a word which is dedicated to one meaning in a context reserved for quite a different meaning is very likely to result in nonsense, and little more needs to be said. But there are problems with this purely meaning-based account. *Request*, for example, *does* correspond to a solicitation of action:

(3) a. I requested assistance from the police.

b. I will request that the police provide me with an escort.

Moreover, in (4), it seems clear that *inquire* is actually being used to solicit a kind of action, rather than information alone:

(4) I inquired whether Robin would be so good as to close the window.

Yet, as noted, for many speakers who find the examples in (3) fine, (1)b is still less than perfect. And while the effect of using *inquire* in (4) in this way is a degree of politeness that borders on unfriendliness in its formality, it is clear that on the most natural interpretation, the speaker is asking Robin to close the

Cambridge University Press
978-1-107-01888-4 — Syntactic Analysis
Robert D. Levine
Excerpt
More Information

window, rather than any far-fetched alternative meaning involving Robin's self-assessment of his or her potential 'goodness.' In effect, one might take the verb *inquire* to be 'coerced,' by certain kind of assumptions about implicit meanings in ordinary discourse, into taking on the sense of a verb of command. For English speakers who display the kind of judgments of the data in (1) already noted, then, something other than simple 'overt' meanings seem to be involved.

Similar observations could be made with respect to *demand*: *I demanded him to leave the room* is judged as awful by virtually all native speakers, in spite of the fact that *demanded* is a solicitation of action roughly comparable in force with *order*, which works perfectly here: *I ordered him to leave the room*. Note, finally, that there is no problem with *demand* as a verb. Again, something besides the notion of meaning seems to be called for.

Examples leading to the same conclusion can be found with little effort in all human languages which have been examined in sufficient detail. A particularly nice illustration for English is given in (5):

(5) a. They charged Robin with perjury.
b. They accused Robin of perjury.
c. They indicted Robin for perjury.

Charge, *accuse*, *indict*, and similar words, which can be characterized as verbs of judicial sanction, are extremely close in meaning. Yet in each of these cases, replacing the preposition in the example with one of the other two will yield a bad result (notated henceforth with a preceding asterisk): **They charged Robin of perjury* is universally regarded as unacceptable, even though, in judicial or quasi-judicial contexts, *charge* and *accuse* are virtually identical in meaning. Another example of this sort is shown in (6):

(6) a. Robin $\left\{\begin{array}{c}\text{grew}\\ \text{became}\end{array}\right\}$ belligerent.
b. Robin $\left\{\begin{array}{c}\text{grew}\\ \text{*became}\end{array}\right\}$ to hate mathematics.
c. Robin $\left\{\begin{array}{c}\text{*grew}\\ \text{became}\end{array}\right\}$ a skeptic about the usefulness of standardized tests.

The property of being belligerent, and the property of being a skeptic, that is, skeptical, both seem to be the same sort of thing; just as being belligerent and hating mathematics both seem to refer to mental/emotional attitudes. So why do the facts sort the way they do in (6)? Indeed, if we replace *a skeptic* with *skeptical* in (6)c, *grew* becomes just as acceptable as in (6)a. Once again, meaning does not appear to be a particularly promising angle from which to attack the problem posed by (6).

Still another instance of the same mismatch between syntactic form and semantic interpretation is afforded by cases such as

(7) We $\left\{\begin{array}{c}\text{solved}\\ \text{resolved}\end{array}\right\}$ the problem.

(8) a. We arrived at a solution $\left\{\begin{matrix}\text{to}\\ \text{??*of}\end{matrix}\right\}$ the problem.
b. We arrived at a resolution $\left\{\begin{matrix}\text{of}\\ \text{??*to}\end{matrix}\right\}$ the problem.

The notation * indicates outright ill-formedness, and ??* a highly marginal status at best. Yet given that we often use *solve* and *resolve* to mean exactly the same thing with respect to problems (particularly of the human conflict sort), a meaning-based view of the conditions under which the prepositions *to* and *of* can appear makes the facts in (8) quite mysterious.

A somewhat more complex example of the same difficulty is afforded by cases like the following:

(9) a. That jail sentence confused the twins.
b. The twins were confused by that jail sentence.

The formal connection between such sentences, which we will investigate in some detail in Chapter 4, is referred to by linguists as the active/passive relationship, and (9) is completely representative of the fact that passives invariably mean the same thing as their corresponding active forms. One very clear pointer to this identity in meaning is that if an active sentence is true, the passive analogue is true and vice versa; we say that the active and passive forms are *truth-conditionally equivalent*. What changes going from the active to the passive is, very roughly speaking, which of two different roles – the initiator of an action vs the target of that action, to take a very common kind of active/passive relationship – gets assigned to which element in the sentence (*the twins* follows the verb in (9)a, but precedes it in (9)b). But the situation depicted by the active and passive is invariably the same.

But things appear to be very different in the case of (10).

(10) a. The twin's imprisonment confused them.
b. They were confused by the twin's imprisonment.
c. In spite of their legal sophistication, the twins were confused at being sentenced to imprisonment.
d. Friends of theirs were confused by the twin's imprisonment.

Native speakers of English are unanimous in their judgment that in sentences such as (10)a, the third person plural pronoun can be taken to refer to the same two people as *the twins*, while in (10)b the pronoun *cannot* refer to those people. In the latter case, *they* must be someone other than the twins. Since, as we have already observed, active and passive sentences have the same meaning, what is responsible for the restriction on who it is that *they* can refer to in (10)b?

Again, there is a seemingly obvious answer based on meaning considerations. Pronouns such as *they* are, after all, pointers to individuals, objects or other types of entity which typically have already been mentioned in the sentence or the discourse, whereas in (10)b, the intended referent – the thing(s) the pronoun is pointing to – is mentioned only *after* the pronoun. How can a linguistic expression intended to refer to some preceding element be successfully used in a sentence to point to an entity which has not yet occurred in that sentence? Hence, the story goes, it makes sense that reference to the twins is ruled out for *they* in this example: one should not be able to use a form to point back to something which has not yet appeared.

But while this answer is plausible, and appealing in its simplicity, it is shown to be incorrect by the data in (10)c–d. We see that while the possessive pronouns *their/theirs* precedes *the twins*, we have no problem interpreting the two expressions as referring to the same individuals in either of these examples. Grasping at straws, one might try to explain the difference in the cases as a reflection of the difference between possessive forms of pronouns, which one might assume can find their referent in a following linguistic expression, versus nonpossessive pronouns which can only share a referent with a preceding expression. But this rather contrived explanation cannot be maintained, as examples such as the following show:

(11) a. Those stories about them are not what the twins wanted to hear.
b. After they were taken to the dean, the twins received a stern lecture about plagiarism and intellectual honesty.
c. Which stories about them do you think the twins would least like to have repeated?

In all cases, *they/them* can refer to the twins.

So it turns out that the simple account of (10) in terms of the relative linear order of the pronoun and its referent fails in the face of quite ordinary and straightforward facts about English, a few of which are given in (10)–(11). For our present purposes, this result means that we cannot take the position either that there is some semantic reason why the possible meanings of the pronouns in (10)a and b respectively are different (after all, actives and their corresponding passive mean the same thing) or that these different possibilities can be accounted for in terms of simple linear order of words within sentences (since whether or not a pronoun can refer to the same thing as an expression such as *the twins* does not seem to be determined by which of these precedes the other).

A final example will help set the stage for the next step in the development of a practical approach to identifying the factors which determine what is and is not possible in the form of any given sentence. Consider:

(12) a. You may want to laugh at them, but you should not laugh at them.
b. You may want to laugh at them, but you should not.

Following the negative marker *not*, we can omit the word string *laugh at them* with no loss of acceptability; in fact, unless fairly heavy stress falls on *not* in (12)a, the latter seems a bit stilted and artificial. Compare (12) with:

(13) a. You may want to laugh at them, but you should try to not laugh at them.
b. *You may want to laugh at them, but you should try to not.

For some reason, the result of omitting *laugh at them* after *to* here yields a very ill-formed example, one which most English speakers find distinctly anomalous. One response to this discrepancy between (12) and (13) is to suppose that the existence of an alternative order which *is* acceptable – *You may want to laugh at them, but you should try not to* – has the effect of making the first ordering of words bad. But this explanation lacks credibility, given that we have the same alternative order available for the well-formed (13)a – *You may want to laugh at them, but you should try not to laugh at them* – with no resulting ill-effects. Both this ordering and the ordering in (13)a are fine. And in general, variant word orders which correspond to identical meanings are not exactly rare in natural languages; note, for example,

(14) a. That Robin is a spy seems strange.
b. It seems strange that Robin is a spy.

(15) a. Robin strode fearlessly into the room.
b. Into the room strode Robin fearlessly.

(16) a. I'm fairly sure that everyone likes ice cream.
b. Ice cream, I'm fairly sure that everyone likes.

In the first of these examples, the only difference in the collection of words used in the two sentences is a meaningless *it*, commonly referred to as a 'dummy' pronoun, which adds precisely nothing to the meaning, since the first sentence conveys exactly the same information without this *it* being present. In the second and third examples, the inventory of words in the two sentences is exactly the same, but the order in the a and b versions is significantly different in both cases. Yet in all three examples, both of the orders displayed in a and b are perfectly acceptable. The rhetorical effect might be slightly different, but the meanings are – in the truth-conditional sense referred to earlier – identical. So whatever the problem with (13)b is, it's very unlikely to be the existence of an alternative order in and of itself.

Such facts, and scores of similar ones that can be easily discovered once one starts looking for them, make it seem reasonable to be very skeptical about the possibility of using either word meanings or the simple linear arrangement of words to explain the somewhat mysterious behavior of English sentences; and investigation of any other well-studied language would lead to the same conclusion. On the other hand, it's important not to go so far as to assume that neither meaning nor linear order *ever* play important roles in determining the status of strings of words in the judgment of native speakers – nothing about the preceding discussion would support that position, and it turns out that it too must be rejected on empirical grounds. The right conclusion is that meaning and linear order cannot account for the behavior of word strings in English – that is, the *syntax* of English – *in general*; and the same holds across the board for all other human languages whose syntax we know something about. Clearly, we have to look in a different direction if we want to make sense of syntactic patterns.

1.1.2 A Possible Line of Solution

The data reviewed in the preceding section suggest that there are other properties of sentences, involving neither meaning nor simple observations about word order, which have important consequences for the possible form of English sentences. But this result requires us to attribute properties to sentences that we have no immediate evidence for. We know how sentences are pronounced, obviously. We know what they mean. And we know what order the words they comprise appear in. But apparently there are properties of sentences which bear strongly on their possible form, but which involve none of the obvious characteristics just mentioned. Something else is necessary, which we have no direct information about at this point. Nonetheless, we can hazard some reasonable guesses. The fact that meaning seems to be largely irrelevant to the behavior we've observed points to the likelihood that some aspect of sentence *form* is responsible for this behavior.

One avenue worth exploring is that words enter into relations with other words, or other groups of words, and these relations are independent of meaning. And since the simple order of words also appears to be irrelevant in the crucial cases where we might want to invoke it, it seems a good bet that at least one of the formal properties we're searching for will have to involve relationships among words which are not immediately 'visible'. This line of reasoning leads to the possibility that what is relevant for solutions to the kinds of problems pointed out earlier may be covert arrangements of lexical items. Earlier, we sought to explain the behavior of pronouns on the basis of certain kinds of overt arrangements – an hypothesis which, as we observed, fails empirically. But we can still entertain the possibility that there are hidden, or more abstract, relations, inaccessible to direct observation, which give rise to the observed behavior.

This kind of reasoning is not unusual in any scientific study of particular phenomena; many people can recall high school physics classes in which a large metal tin, the kind commonly used for olive oil, is filled with water and then has the water siphoned out, with the can collapsing into a crushed heap as the water is removed. If this experiment were conducted in a vacuum, the can would remain intact under the same conditions, yet there would be no visible differences in the two situations that would account for the different effects. We would then be inclined to posit some invisible factor responsible, and look for further evidence to challenge or corroborate this hypothesis, eventually leading us to conclude that an invisible medium is indeed present in the context in which the can is crushed, that that same medium allows a flame to be struck from a match (as opposed to the situation in which the can remains intact, and where no match would ignite regardless of how many times it was struck), and so on. Since we're looking at sequences of words, rather than physical objects, however, it's not immediately apparent what linguistic analogue would correspond to the presence vs absence of the invisible gas we call air.

A more specific sort of analogy which may help make this possibility more concrete can be found instead by comparing the sentences in the examples given above with strings of beads that make up different necklaces or bracelets, with each bead corresponding to a word. Imagine being in the position of someone looking at a necklace consisting of a line of beads on a string and wondering if there were any properties of that necklace which couldn't be explained by the simple linear order in which the beads had been set. An obvious candidate for such a concealed property would be in the way the beads are grouped: is there any possibility that what looks like a single uninterrupted strand actually has small knots separating certain groups of beads from others? If there are knots between the fifth bead and the sixth, and betweeth the fifteenth and the sixteenth, then the effect of cutting the string in the exact middle of the necklace is going to have a rather different effect when the two separate halves are picked up by the uncut ends than a necklace with no internal partitions. The same may well be true for two otherwise identical necklaces with knot partitions established in different places respectively.

These examples suggest a way in which seemingly identical strings of elements of the same kind can indirectly reflect hidden differences which are strictly independent of the overt order of the elements that the string comprises. Without actually proving that human language sentences necessarily incorporate structural differences of just this kind, they offer a kind of existence proof that where systematic differences exist in the patterns of behavior manifested by different-looking sentences, the divergence could in principle reflect distinct ways that the words in the two sentences 'cluster,' in some sense still to be made more specific. But since deductive reasoning can only take us so far, we need to turn to some actual data, some phenomenon or set of phenomena which might shed light on the questions we've posed.

1.2 Structure: Some Clues

Our chief tool, here and in the following chapters, will turn out to be the relatively simple idea that where we have sets of sentences which seem to be related to each other in a systematic way, the precise form of the relationship must – wherever a simple account in terms of meaning cannot be given – be accounted for in terms of some *formal* property of the sentences involved.

1.2.1 Displacement: A Basic Natural Language Pattern

We've already mentioned, in passing, one of the key components of the data we need to use as a basis for probing the possibility of hidden linguistic structure: the existence of sentences which mean the same thing, use identical or

CAMBRIDGE

Cambridge University Press
978-1-107-01888-4 — Syntactic Analysis
Robert D. Levine
Excerpt
More Information

nearly identical vocabulary, but manifest significant differences in word order. Consider for example sentences of the form in (17):

(17) a. That Robin turned out to be a spy $\left\{\begin{array}{c}\text{worries}\\ \text{troubles}\\ \text{encourages}\\ \text{annoys}\\ \text{enrages}\\ \text{impresses}\end{array}\right\}$ me.

b. It $\left\{\begin{array}{c}\text{worries}\\ \text{troubles}\\ \text{encourages}\\ \text{annoys}\\ \text{enrages}\\ \text{impresses}\end{array}\right\}$ me that Robin turned out to be a spy.

c. I'm $\left\{\begin{array}{c}\text{worried}\\ \text{troubled}\\ \text{encouraged}\\ \text{annoyed}\\ \text{enraged}\\ \text{impressed}\end{array}\right\}$ that Robin turned out to be a spy.

Not only are these sententences truth-conditionally identical in meaning, but they reflect an orderly and systematic pattern involving near-identical vocabulary arranged in three distinct word-order patterns which relate to each other in a completely regular manner. Furthermore, the same relationship holds if we replace *That Robin turned out to be a spy* with *That my cousin's wife won a Nobel Prize in physics*, or with *That the Dean of the Medical School never pays any taxes*, and so on, and replace *me/I* with *them/they*, or with *the trustees*, or with *an old friend of mine from school*, etc. We can almost always do this sort of thing: record a number of English sentences at random, and then on the basis of these sentences, construct other sentences using almost exactly the same words, which mean exactly the same thing as the original sentences, but nonetheless differ noticeably so far as word order is concerned. We might note, for example, that forms of sentences such as (18)a are systematically matched with sentences such as (18)b:

(18) a. If Terry SAID you have made a mistake, then you HAVE **made a mistake**.
b. If Terry SAID you have made a mistake, then **made a mistake** you (definitely) HAVE __.

In the second of these, the string of words *made a mistake* appears to be missing from what we believe its normal place to be as reflected in the first example; it appears instead at the front of the second part of the larger sentence, directly following *then*. The meanings of the two sentences in (18) are identical, and exactly the same words are present in (18)a and b; all that has changed between the two is the order of a particular substring shared by both sentences.

It is not difficult to find any number of completely parallel examples, for example:

(19) a. Terry SAID she may go to the movies, and indeed, she MAY **go to the movies**.
b. Terry SAID she may go to the movies, and indeed, **go to the movies** she MAY __.

(20) a. If Terry SAID that Robin has been giving Leslie a hard time, then Robin definitely HAS been **giving Leslie a hard time**.
b. If Terry SAID that Robin has been giving Leslie a hard time, then **giving Leslie a hard time** Robin definitely HAS been __.

What are we to make of such examples? The very simplest assumption is that, where we have a series of words that we recognize as a sentence, we also expect to find a different series of words, related to the first by repositioning *any* substring *X* so that *X* appears at the front of the sentence. But this summary of the pattern exhibited in (18)–(20) fares badly in the face of examples such as the following:

(21) a. Terry says she is putting the book on the table, and **putting the book on the table** she is __.
b. *Terry said she is putting the book on the table, and **putting** she is __the book on the table.
c. *Terry said she is putting the book on the table, and **putting the book** she is __on the table.
d. *Terry said she is putting the book on the table, and **putting the book on** she is __the table.

Apparently, not all subsequences of words within a sentence are created equal. There is a clear difference in status between *putting the book on the table* in (21)a and *putting*, *putting the book*, and *putting the book on*. *X* cannot be any arbitrary substring, and to advance our current line of analysis, we need a way to mark the subsequences of words within any give sentence which are displaceable, in contrast to those which are not.

We might start by using some fairly transparent notation, such as brackets ([ ]) around the sequences we've been able to relocate, indicating that these sequences constitute some kind of covert unit. For the sentence *She is putting the book on the table*, we could use this notation to capture the pattern exhibited in (21) as follows:

(22) she is [putting the book on the table]

We further assume that only such unitary strings are able to appear at the front of the sentence, leaving an apparent gap corresponding to the position they occupy in (22). Simple as this move appears to be, it leads immediately to some interesting problems which help us to clarify just what kind of solution it is we're considering.