The Divine Comedy

Foundational Issues in Linguistic Change and Reconstruction

2. Secunda pars eiusdem, que dicitur Purgatorium.

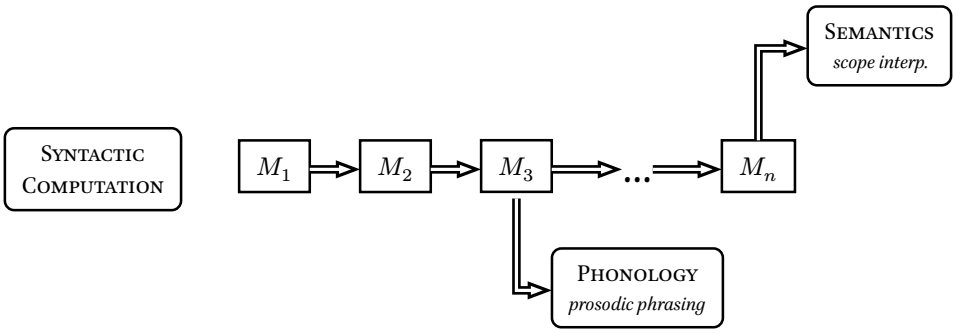
2.1. A Ladder for Ascension: grounding assumptions regarding syntax

1. I hope that by going painfully slowly through the phonological change discussion we've seen just how useful it can be to be as explicit as possible about what foundational concepts such as 'language', 'change', 'phonology' and 'phonetics' are. Let's try to do the same for syntax before we delve into this domain.

2. Given the strong mentalist position we have sketched out above, it is not surprising that I believe that syntax (and, indeed, linguistic structure generally) is one of the computational capacities of the human mind, provided to us by our genetic endowment (like our other properties *qua* humans are). What are the computational capacities of this system, and over what kinds of objects are those operations defined?

3. I will assume that sentences are *constructed* by the syntax out of entities (bundles of features) in the mental lexicon. By 'bundles of features' I just mean 'entities made up of a set of properties' (and you can't be an entity unless you have at least one property, and you can't have precisely the same properties as another entity and count as a different entity). We do not at present have terribly good theories of what the precisely correct set of features might be, though a good number, posited already in antiquity, have proven scientifically productive over a long span of time (noun, verb, etc.).

4. A pretty standard model for some time now has been one in which syntax mediates between sound and meaning, i.e., between phonological and semantic interpretation. Both phonological computation and semantic computation get as their input a syntactic representation (though not necessarily the same syntactic representation). That model might look something like this:



5. A sentence is not just an audible set of lexical items (the latter being how I will refer to the feature bundles from now on). The set {Bill, John, saw}, were they lexical items (they are not, as we will see) could be combined into the audible strings 'John saw Bill', 'Bill saw John', 'Bill, John saw', and 'John, Bill saw' all of which are grammatical for me, and no two of which mean the same thing. The syntax must therefore do more than just *combine* the lexical items. It needs to combine them differently to express different meanings.

6. The strongest hypothesis is that whenever an audible elements are combined differently, they mean something different, and whenever their combination conveys a difference in meaning, they were combined differently. We can then motivate the difference in composition: the elements were combined differently *in order to* express a different meaning. If we do not adopt this strong position, then it is hard to see what factors would cause elements to be combined differently.

7. And again the most restrictive hypothesis is going to be that the syntactic computation is deterministic — given the same set of inputs, the syntax will compute over them in the same way. This entails that if the audible elements are combined differently, the computational system is operating over a different set of inputs. How is this going to work?

8. Let's approach this question from the broader persective of to what extent we want syntax to correlate with semantics? One can imagine three possible relationships between a meaning we believe to be present in a sentence and the syntax of that sentence (for right now we will use 'syntax' to mean placement in the [surface] string—a definition that we will revise later on).

    - **Type a:** there is a clear and unambiguous relationship between the syntactic distribution of an element and the semantics we are interested in.
    - **Type b:** It is not clear to us what the relationship between the syntactic distribution of an element and the semantics we are interested in.
    - **Type c:** It is clear that there is no relationship between the syntactic distribution of an element and the semantics we are interested in.

9. Suppose I say 'I kicked an elephant'. In the correct analysis of the meaning of what I have said, one will need to make reference to some grayish kind of mammal with a big trunk (which is about the totality of my knowledge of elephants, and I'm kind of guessing about the 'gray' and 'mammal' parts). This meaning will be present in any sentence I say involving the word 'elephant'. In particular, the presence of that semantics in no way depends on whether the word 'elephant' is the object of a transitive verb (as in the example), or the subject of a transitive verb, or the subject of an intransitive verb, or the object of a preposition, the possessor of another noun, etc., etc. Clearly the meaning 'grayish mammal with a long trunk' is *not* encoded by the syntax: it persists under any licit syntactic perturbation of the string, as long as that string contains the lexical item 'elephant'. Since this meaning is tied to the lexeme 'elephant', let us call it *lexical semantics*. It is the lexical semantics of elements that the syntax is combining: 'combining' doesn't eliminate the meanings of the elements combined.

10. So there is a kind of semantics present in the clause for which syntax is demonstrably not responsible (i.e., we have at least one case of Type (c) above.

11. There are also clear examples where, in every language I have ever seen, there appears to be evidence of a cross-linguistically necessary relationship between the syntactic distribution of an element and its semantic interpretation. If we take a word like 'and' or 'or', that word will bring its lexical semantics to the clause (like 'elephant' does), 'and' being 'coordination' and 'or' being 'disjunction', but, in addition, in all the languages I know, these elements must appear *adjacent to* an element being conjoined or disjoined from another element, rather than in some arbitrary (or arbitrarily fixed) position in their clause. Conjunctions such as this differ from 'elephant' in that they take *arguments*—elements over which their semantics operate. Their argument structure must be *saturated* (i.e., satisfied by syntactic combination with some appropriate type of entity) for the string to be well-formed (i.e., coherent, in the technical sense).

12. We thus have clear instances of syntax-semantic relationships of Type (a) above.

13. But the evidence for Type (a) is much stronger than that example implies, and in trivial ways that we are so used to, it is easy to take them for granted. For example, if we take a sentence like: *when the big elephant saw a small duck, this brown cow mooed* the elements involved in the 'seeing' event are 'the big elephant' and 'a small duck'—'this brown cow' is in no way claimed to be involved. By the same token, it is 'this brown cow' that took part in the 'mooing', the elephant and duck have nothing to do with it. This *clause-boundedness* of semantic interpretation represents another cross-linguistically trivial case of there being a firm and well-defined relationship between syntactic structure and semantic interpretation. We can make sense of it fairly trivially is we assume that (1) the syntax builds clauses and (2) semantic interpretation (excluding lexical semantics) is a function of the syntactic construction process.

14. Moreover, it is the elephant that is claimed to be 'big', the duck 'small', and the cow 'brown'. We are dealing with a specific big elephant, an indefinite small duck, and a near-the-speaker brown cow. The relationships established within the Noun Phrases we see here are all both syntactically and semantically *local*. This makes sense, again, if we build the major elements out of which clauses are constructed (noun phrases and verb phrases) in the syntax, and adopt the stance that semantic interpretation above the level of lexical semantics is a function of the syntactic structures being constructed.

15. Of course this is all well-known, and was codified quite clearly by Behaghel, as his 'first law' (Behaghel 1932 §1426):[1]

> Das oberste Gesetz ist dieses, daß das geistig eng Zusammengehörige auch eng zusammengestellt wird.

[1]'The foremost law is this: that that which goes together closely psychically is also placed closely together.'

16. This is the central notion, to this day, of *constituency*. All human languages we have ever observed provide evidence for constituency (clauses, phrases, etc.). That constituency is the *result* of syntactic concatenation: if the syntax combines two elements, they are part of the same constituent. There can be no doubt that *some* meanings are encoded syntactically in this sense. In addition, there is in my view unambiguous evidence that some meanings (the fact that clauses containing the word 'elephant' have, in their meaning, some kind of reference to mammals with trunks) are *not* encoded syntactically.

17. What about the cases in which we can't tell whether the relevant meanings are encoded syntactically? How should we approach these? Again, let's work from a simple example: *Your theory, any scholar who has half a brain would really hate!*

18. Included in the semantic interpretation of this sentence is the fact that the thing that is not liked by the subject ('any scholar who has half a brain') is 'your theory'. That fact does not seem to be syntactically encoded because the entity that is 'hated' in a sentence normally appears within the verb phrase constituent whose head is 'hate'. 'Your theory' in the sentence above seems pretty clearly *not* to be within the verb phrase headed by 'hate' – it is not *local* to hate.

19. Note that this seems like a blatant violation of Behaghel's First Law: 'hate' and 'your theory' belong closely together *psychically*, but they are not bound closely to one another *syntactically*. If we take Behaghel's First Law seriously as a *law*, then we need to conceive of apparent violations of that law as just that—merely apparent. In essence, this involves the making of a relatively simple (Occam!) assumption: the meanings encoded in a sentence arise in one of two ways, either

    - they come from the lexicon, where they are stored together with phonological and morphosyntactic information about the morphemes the speaker knows; or,
    - they result from the relationships established by the syntax.

20. Recall how Occam works: if it is possible to construct a model that accounts for the relevant aspects of the meaning of a clause using just these two processes, one cannot add another (e.g., 'search around the local domain — in some sense of that word — to find things to construe [=interpret] together, regardless of their syntactic relationship').

21. Indeed, in the most restrictive interpretation of things, the syntax cannot *add* meaning at all. Instead, as in the 'and' case, it simply *links* (via combination) the lexical semantics of a given lexical item into an *argument slot* of another lexical item. All *meanings* come in on lexical items, and all *meaning relationships* are established by bringing lexical items into a structural relationship via concatenation. Like all theories, this could easily be wrong, of course, but it does have the virtue of *providing an explicit analysis* of any sentence for which we have an interpretation. Let's return to the theory-hating example and see how.

22. We have an *apparent* violation of Behaghel, in that the direct-object semantics which should be expressed by filling the object position of 'hate' by 'your theory' appears to have never been created via concatenation, yet that meaning is present. Since the meaning is present, under the assumptions posited above it *must* have been created. So the sentence must have, at some point in its derivation, looked like this: *Any scholar who has half a brain would really hate your theory!*

23. But that isn't the sentence we are interested in, and it doesn't have the meaning our original sentence has, because 'your theory' is strongly focussed in the original sentence, but is not so focussed in our structure. How is this 'strong contastive focus' semantics expressed? Recall, the syntax cannot *add* meaning, so we cannot say that the syntax simply 'fronts a direct object to give it strong contrastive focus'. We need to get the FOCUS semantics from the lexicon and put it into the syntax, if we are going to have it be present in the output. This approach has some disadvantages, but many, many advantages.

24. The big disadvantage is that we need *phonologically null morphemes*.[2] I personally don't consider this any real disadvantage (I assume plural 'sheep' has a null plural suffix, e.g.), but it really bothers some people. My advice would be that you not let it bother you, and your life will be much easier, syntactically speaking, but you can do what you want.

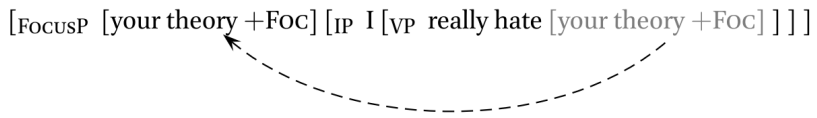
25. The advantages include the fact that because it is drawn from the lexicon it will be placed in the tree in some well-defined position: this will be the position to which 'your theory' moves when it is satisfies an argument slot that FOCUS brings into the tree. This works much better than the 'front the direct object to focus it' analysis, because 'front' is ill-defined:

(1) a. Bill's theory is terrible, but I really hate **your theory**.
b. *__Your theory__ Bill's theory is terrible but I really hate.
c. *Bill's theory is terrible, **your theory** but I really hate.
d. Bill's theory is terrible, but **your theory** I really hate.

(2) a. I don't think much of you and I hereby declare that I really hate **your theory**.
b. *__Your theory__ I don't think much of you and I hereby declare that I really hate.
c. *I don't think much of you **your theory** and I hereby declare that I really hate.
d. *I don't think much of you and I hereby declare **your theory** that I really hate.
e. I don't think much of you and I hereby declare that **your theory** I really hate.

[2] Actually, in this case (but not in some others), there is probably special prosody associated with the strong contrastive focus, so the morpheme is only *segmentally empty*, while it almost certainly has suprasegmental content.

26. The 'landing site' for such fronted objects appears to be some well-defined structural position, not 'the front' of anything obvious. Indeed, we can define an entity that this can be the front of, but if syntactic elements arise via syntactic concatenation (again, the simplest theory), VPs being the projection of Vs, NPs being projections of Ns, etc., then we probably want the type of clause with a FOCUS on its left to be a FOCUSP, and to be a projection of FOCUS. FOCUS is that phonologically-null lexical item that is inserted in the tree in the place strongly contrastive focus objects move to! (I put a +FOC marker on the element to focussed, since otherwise the syntax wouldn't know what to move! We don't want it focussing things at random!)



27. What about Behaghel? Well, what we see is not that Behaghel is being *violated*, but that the syntax is absolutely abiding by the requirements of Behagel's First Law very, very strictly. Since [your theory +FOC] is an argument of 'hate', it *must* be 'zusammengestellt' with 'hate'. And it is. And, since it is the argument of the FOCUS operator that heads the phrase, it *must*, by Behagel's Law, get 'zusammengestellt' with that entity. It thus occupies two positions in the tree in the derivation.

28. It is this relationship between two structural positions that the traditional generative syntactic notion of 'movement' was a metaphor for. The theory never involved actual 'movement', of course (or else one could have asked 'how fast'?): it was always about structural relations.

29. If we, finally, stipulate that only the highest version of the twice-appearing element in the tree above gets pronounced, we can see how the pronunciation of the clause is derived. The reason why it is hard to see how the semantics and syntax are related in this string is simple, then: only some aspects of the syntactic representation get pronounced. If we examine the tree, as it emerges from the syntax, the syntax-semantics relationship follows from a very simple theory. It is only from the fact that pronunciation fails to give physical realization to all aspects of that tree structure that unclarity arises.

30. And we already knew that we had this problem. In ambiguous sentences like 'The zookeeper poked the gorilla with the banana' it is clear that the *tree* created by the speaker is *not ambiguous*. Only the *acoustic string*, stripped of the hierarchical information embedded in the tree, is ambiguous. It is not surprising that when one strips information out of a representation, ambiguities about its true structure arise.

31. Okay, now we recognize that Behagel's First Law is a *law* — it cannot be violated. We can now ask the question of what meanings are 'lexical' (and thus irrelevant to the syntax, like the greyness of elephants) and which ones are 'syntactically relevant'. It turns out that the construction of clauses involves the construction of at least three quite distinct *kinds* of 'syntactically-relevant' meaning:

    - CORE PREDICATION: who did what to whom? (argument structure; the 'VP domain')

- Subjecthood & Modality: what is the subject of the sentence, and under what modality does the predication hold of the subject? (tense, aspect, mood, negation, etc.; the 'IP domain')
- Information Structuring: how does this predication relate to the broader discourse context, including 'discourse-related' functions such as 'Topic' and 'Focus', as well as clause-typing (interrogative vs. declarative, etc.; the 'CP domain')

32. To get a better grip on the concept of 'syntactically-relevant' meaning, let me say a little about how we might want to define what a 'sentence' is, because it probably isn't exactly what you're thinking a sentence is. Recall that we want to define our computational systems as operating over a set of entities, and that the most efficient way to rule out impossible syntactic systems is to not give them computational access to the entities they don't seem to be able to compute over (rather than by stipulating in a list everything they can't do).

33. It seems clear enough that there are aspects of each lexical item that the syntax has access to (it treats some lexical items differently than others) and other aspects of those lexical items that the phonology (e.g.) has access to. There can never be *literal* wh-movement, under which words that start with 'wh-' (i.e., [w-] in my dialect) undergo a syntactic displacement. We block this impossible human language by saying that the syntax cannot 'see' what the phonological segments in the entity it is computing over are, and thus cannot manipulate objects based on those segments.

34. Likewise, the phonology cannot establish the scope relations between 'and' and the element it gets combined with ('and Mary' in 'John and Mary'), because it doesn't know that 'and' is a coordinator, or how many arguments it takes, or whether 'Mary' has the appropriate properties to satisfy the argument structure of 'and' etc.

35. This is basic modularity. But note that it arises not from an *a priori* stipulation about the structure of the grammar, but from the basic scientific orientation of our undertaking. We have a 'syntax' module that is distinct from a 'phonology' module because we find two distinct sets of lexical properties being accessed by the 'syntax' and the 'phonology', respectively. Combining the two domains into a single system both *allows* and, therefore, *predicts the existence of* whole classes of grammars that we never see in the world.

36. There is also no human syntax in which words which refer to 'ducks and duck-like beings' must undergo syntactic operations because of their duck-referring properties. Nouns may fall into various classes and may be manipulated based on those class memberships (grammatical gender or noun class membership, mass vs. count etc.) by the syntax, but duckhood isn't one of the possible classes. How do we make sure there is no 'duck-movement' (not actual ducks, obviously…)? We preclude the syntax from knowing that the element it is operating on is a [+duck] word.

37. The same is true of cows, so: 'This is a cow.' and 'This is a duck.' differ only in ways that (in languages without arbitrary grammatical gender, like mine) are invisible to the syntax—i.e., they don't differ for the syntax. If they don't differ for the syntax then, *qua* syntactic objects, they are the same. Since if I say them I don't say the same thing, nor mean the same thing, we need to distinguish between syntax-internal considerations (under which these are the same thing) and all the other stuff we have to worry about in discussing language (under which they differ). Since I am inclined to believe that syntax is about sentences, I like to call the *one object* that these two phonological strings ('this is a cow' vs. 'this is a duck') instantiate a **sentence**. So, two strings, two phonologies, two meanings, **one sentence**.

38. Given these considerations, I will assume that the syntax has as a general property no access to what we called above 'lexical semantics', i.e., to the 'encyclopedic' content of lexical items (e.g., whatever makes 'hat' different from 'jacket' or 'chair') nor, of course, to their phonological content (hence no 'hat'-fronting, or 'word-with-*b*-in-it rightward extraposition). Thus, assuming identity of all syntactically-relevant features (grammatical gender, grammatical animacy, etc.), 'the cats ate this bread' is *the same sentence* as 'those maruts drank the soma'. The two utterances generated from these sentences differ in their phonology (because of the phonological features of their lexical matter, not visible to the syntax, and thus not part of the syntactic computation or output) and in some aspects of their interpretation (because of the encyclopedic semantics of the lexical items, fed into the conceptual interpretive system), but not in their syntax (nor in their syntactic representation).

39. Note that this means that the pronunciation of sentences is both too poor to actually reflect what is going on (as we argued above about focus structures and 'the zookeeper poked the gorilla with the banana') and too rich to accurately reflex the syntax as such (because it differentiates between 'the cats ate this bread' and 'those maruts drank the soma').

40. This should not be surprising: it is the normal situation when a scientist gets data from the world. If you look at a table, you both *fail to get* information about its true structure (a set of subatomic particles clumped in various ways in what is mostly empty space) and *get* all kinds of (mis)information that misleads you as to its true structure (it's solid, it's stationary, etc.). The true world is both hidden from us **and** deceptive in how is manifests itself to us. Science is about transcending this deception and uncovering the hidden. That is why it is hard.

41. At any rate, an interesting observation arises from the theory that the syntax builds up three different kinds of meaning in the course of the derivation (as outlined in above)—it appears that the syntactic structure is constructed sequentially (first the Core Predication, then the Modality, then the Information Structuring). An implication of this observation is that the *accessibility* of each of these domains for analysis differs quite dramatically in the semantic vs. syntactic areas.

42. Because building higher-order structure *preserves* the semantic relations established by the Core Predication (the patient remains a patient even when externalized into IP under passivization, and a direct object keeps its thematic role even if topicalized or focused), and because we have millenia of experience working with notions like 'agent' and 'patient', 'locative' and 'dative', etc., it is generally quite straightforward to figure out the semantics of the core syntactic relations for a clause, if we are dealing with a language that we know at all well.

43. By contrast, our theories of the semantics of tense, mood, aspect, and the like (i.e., IP-domain functions) are much more weakly articulated, and these categories are much harder to establish (think of the wealth of 19th-century 'Indo-European syntax' literature that attempts to sketch out the conditions on the use of the subjunctive or optative in archaic IE languages).

44. And at the CP-domain level, a firm understanding of even what the basic categories are, and what they mean, is still at the early stages of development: what does it mean to say something is a 'topic' or 'in focus'? Does it mean one thing, or are there many different kinds of 'topic' and 'focus' elements?

45. Therefore, *semantically,* i.e., in terms of interpretation, we get a pattern of scientific accessibility that goes (from easiest to most challenging)

Core Predication > Modality > Information Structuring.

46. By sharp contrast, syntactic *structure* (as opposed to syntactic meaning) is *not* audibly preserved as we move through the stages of the derivation. Syntactic relationships established in the Core Predication can easily fail to have any direct physical manifestation in the output, elements having been displaced into both (or either of) the IP-domain and the CP-domain.

47. Thus, in terms of *syntactic structure* (rather than semantic interpretation), the domain constructed *last* remains the most accessible to us, while that constructed first has had the greatest opportunity for disruption (and it thus the hardest to determine the properties of). That is, we get a pattern of scientific accessibility that goes (from easiest to most challenging)

Information Structuring > Modality > Core Predication.

48. It follows, somewhat unfortunately, that we will have the easiest time, as we develop our understanding of, e.g., ancient texts, in figuring out 'who did what to whom'. It is, indeed, generally this level of comprehension that we have richly developed for Vedic Sanskrit or Attic Greek. I say 'somewhat unfortunately' because the *syntactic level* at which the semantic relations involved in this interpretation are established remains the hardest for us to get a grip on, since it is the domain most likely to be disrupted by subsequent operations.

49. Just one final note on the three domains (VP, IP, CP) of syntactic construction: it does not appear that there's much good evidence for movement *within* these domains. Entities don't 'move' from the position of the patient complement of a verb (within the VP) to the position of a recipient of that same verb (probably also within the VP), e.g., or from 'topic' position to 'focus' position (in the CP domain). If this is a general constraint, it will have interesting implications for our analysis of 'discontinuous constituency' in archaic Indo-European languages.

50. And one very final point on syntax generally. What gives rise to seeming diversity of syntactic systems? Why isn't English syntax exactly like German, or Japanese, or Attic Greek syntax? One way, perhaps now familiar to you, to explore this issue is to ask what the acquirer needs to learn to end up with an English-type syntax, or a Japanese-type syntax.

51. Recall what we believe the syntax is: it is a computational system (the syntax) which computes over the morphosyntactic properties (features) of a set of lexical items (the lexicon).[3] Under this conception of things, there are three possible ways in which English syntax could look different than Japanese syntax, or, to put it in a diachronic context, in which Attic Greek syntax could look different from Proto-Indo-European syntax:

    - Attic Greek and Proto-Indo-European could differ in the morpho-syntactic properties on their lexical items (and/or on their inventory of lexical items).
    - Attic Greek and Proto-Indo-European could differ in the kinds of syntactic computations they perform on (the same) lexical items.
    - Attic Greek and Proto-Indo-European could differ both in the morpho-syntactic properties of their lexical items and in the types of syntactic computations they perform over those morphosyntactic properties.

52. I like to use a cookie-making machine analogy to discuss this, so here goes. Imagine you have a black box which generates cookies out of one end after you pour ingredients into the other. Our three scenarios for explaining why you sometimes get chocolate chip cookies out of the machine (Attic Greek) and sometimes get oatmeal cookies (Proto-Indo-European) now look like this:

    - You get the difference in cookie outputs based on the ingredients you put into the machine, which identifies the ingredients and treats them always in the same manner.
    - You get the difference in cookie outputs even though you put the same ingredients in. It's a black box, so you don't know why, but somehow it computes differently over identical objects.
    - You put different ingredients in, and the machine produces two kinds of cookie, but the kinds of a cookie are not a direct function of the ingredients you put (that would

[3] If you believe that the syntax and lexicon together do not determine the (relevant aspects of) the semantic interpretation of the clause, then you would also need a 'semantic construal system'. In my view, this will end up looking exactly like the representations posited in our 'syntactic movement' system, and are thus completely irrelevant. To the extent they do not reduce to 'syntactic movement', I do not personally see how anyone could ever learn them.

be theory (1)), but, as in theory (2), sometimes the machine computes differently over the same ingredients.

53. You are used to hearing me invoke Occam's Razor by now, and it is relevant here was well. In my view, all but the first cookie story looks pretty crazy, and like a very bad theory of what the 'black box' cookie-making machine is doing. The most restrictive theory is one that ties the input directly to the output, rather than adding some mysterious non-deterministic operations within the box.

54. Also, in the grammar (rather than cookie) case, we can ask the acquisition question. That question looks like this: of the two things involved, either (or both) of which *could* be responsible for variety in output, do we believe with a high degree of confidence that one of them must be learned? If so, the simplest theory is clearly going to be that the variation is located in that domain (since we know learning of grammatical structures involves imperfect, noisy transmission, and we believe that such noisy transmission leads to the kind of re-analysis that we believe 'change' is).

55. It seems obvious (and is not disputed, to my knowledge) that the lexicon must be learned by the acquirer. For example, the fact that the meaning 'and' is phonologically /ænd/ in English and that that phonological form is linked to a syntactic feature that creates a structure which serves as a barrier to extraction,[4] but that it is phonologically /kaj/ in Greek and does not create such a barrier to extraction, can hardly be due to some people having 'Attic Greek-type genes' and others 'English-type genes', thus it must be learned (if known, and not innate).

56. A single black-box grammar, operating on an Attic Greek-type lexicon could, in principle, produce Attic Greek-looking output, but operating on a Proto-Indo-European-type lexicon, it could be expected to produce Proto-Indo-European-type output. Occam Razor requires, if we are to assume that in addition to this *agreed upon* and *conceptually necessary* lexical variation, which already predicts differences in output, we are going to posit *additional machinery*, e.g., differences in the nature of syntactic computation, to account for differences in output, that there be compelling evidence that that additional machinery is also needed.

57. And in the presence of a varying lexicon (i.e., the possibility of different inputs), how could the speaker figure out differences in syntactic computation? Let us return to the delicious cookie metaphor. Imagine you have a black box. You push a button on the front and out comes a chocolate chip cookie. You push the button again and out comes an oatmeal cookie. You make the theory that things like flour, eggs, sugar, butter and chocolate chips get put into the machine in the situation when it produces chocolate chip cookies, and flour, eggs, sugar, butter and oatmeal get put into the machine when it produces oatmeal cookies. It would be completely illict, as a scientist, to *additionally* assume that there were two different machines in the black box! The difference in ingredients provides all the explanatory power you need.

[4] The well-known 'coordinate structure constraint' of early generative work, designed to allow 'who_i did you see at the party who_i?' but to block 'who_i did you see who_i and John at the party?'

58. So, the syntactic component of the human mind puts things together, creating semantic relationships (the only semantics in the clause that is not given by the lexical semantics of the items computed over). It creates relationships based on the properties of the entities it is computing over, so it does not treat the noun 'duck' the same as the conjunction 'and' (because they differ in their morphosyntactic properties). Different languages have different lexical items, which have been constructed (i.e., assigned morphosyntactic and phonological features) in the course of acquisition. Computation over different objects gives different outputs. The most restrictive theory is one in which this is the only source for cross-linguistic variation in syntactic structure (the Borer-Chomsky Hypothesis, as it is now often called).

59. Synchronic cross-linguistic syntactic variation is just one manifestation of the differences that can exist between human linguistic systems; diachronic change is another. If synchronic cross-linguistic syntactic variation is due to differences in the input to the computation, syntactic change is to be explained with reference to such differences as well (Hale 1998).

2.2. Wackernagel's So-called Law: Clitics, Prosody, Syntax

60. Wackernagel's Law (Wackernagel 1892) is a generalization about the surface distribution of enclitic elements in a variety of archaic Indo-European languages, and thus, arguably, about Proto-Indo-European itself. It is stated quite clearly in Wackernagel (1920:7) — I have inserted some underlying for key conceptual issues we will discuss momentarily:[5]

> Z. B. im ältesten Griechisch, in sehr hohem Masse bei Homer, auch noch bei Herodot, ist das Gesetz lebendig, dass schwach betonte Wörtchen, welches immer ihre syntaktische Beziehung sei, unmittelbar hinter das erste Wort des Satzes gestellt werden.

61. There are two reasons why this 'law' should figure prominently in any serious discussion of the syntax of Indo-European, and no doubt of diachronic syntax generally. First, it is widely praised as a foundational concept in IE, and general diachronic syntax, e.g.:

> One of the few generally accepted syntactic statements about IE is Wackernagel's Law, that enclitics originally occupied the second position in the sentence. (Watkins (1964:1036)

> Wackernagel's Law, that clitics tend to appear in sentence second position, is one of the firmest discoveries in the history of syntactic change… (Harris & Campbell 1995:24)

[5] 'For example, in the most archaic Greek, to a very high degree in Homer, also still in Herodotus, the law is active that weakly stressed little words, whatever their syntactic relationships might be, are placed directly after the first word of their clause.'

62. Second, it touches upon all of the foundational issues for the field which we have been dealing with in up to now: constituency, the interactions of the modules (particularly the phonology and the syntax), textual interpretation and the nature of the mind! If we can make some progress on understanding the nature of 'Wackernagel' phenomena, we must be doing something right regarding these fundamental matters. If not, we are probably not.

63. As we noted earlier, a good way to understand a system is to ask oneself what is the set of *entities* (i.e., properties) posited by the system and what is the set of processes that are defined over those entities. The interaction of the posited entities and processes should give rise to the desired observational behavior.

64. Given the importance of Wackernagel's Law for the study of historical syntax, it is worth asking ourselves about it in the first instance from Wackernagel's own perspective. What kinds of entities and processes did he envision in his original paper? We won't deal with *all* aspects of his model (that would take too long), but nothing too crucial will be left out.

65. Let's start with the set of entities Wackernagel seems to posit, explicitly or implicitly.

    - There is a set of 'unstressed' elements to which the Law applies. They are enclitics, and we can designate them as members of the class **cl**.
    - There is a set of elements which bear an accent mark, but which otherwise seem to behave like the clitic class. Wackernagel called these 'postpositives', so we'll designate them as **Ṕ**.[6]
    - There is a set of elements that *don't count* (or, to be a bit more precise, do not have to count) for determining second position—much as German *aber* does not count for the establishment of 'verb second' in that language. These have come to be called 'null position' elements in more recent literature. For Wackernagel, Greek *allá* 'but' is an example. Membership in this class must be stipulated, as near as can be determined from the discussion. We will designate these elements as ∅.
    - There is, of course, a set of elements which definitely *do* count as occupy first position — these are typically open class lexical items; they are stressed, of course. We will designate these **X́**.

66. As far as the processes which give rise to Wackernagel's Law placement of clitics, Wackernagel himself seemed to have a 'templatic' approach in mind. The idea of such approaches is that there is an (abstract) set of 'slots', arranged in some order, into which syntactic elements are to be positioned. Syntactic operations consist, under such a conception, of placing elements into their appropriate slots. We can thus describe the structure of the clause by specifying the abstract arrangement of these slots.

67. For sentences with at least one clitic, the left edge of the template seems to look like something like this:

[6]The Greek sentence connective *dé* is in this class, but in general membership in this class is somewhat difficult to define: there is a partial definition (you must bear an accent), and presumably you must not be an open-class elements (such as a noun or non-copular verb) but how precisely you are to be distinguished from other functional elements *aside from your syntax* (which of course is what we are trying to explain by creating a class for you) is tricky.

$\varnothing$ $\acute{\mathbf{X}}_1$ **Ṕ cl** $\acute{\mathbf{X}}_2$ …

68. Note that I have left in the above to one side the issue of how to get the 'template' to allow multiple **Ṕ** and **cl** elements, but at the same time to *not* allow for multiple $\acute{\mathbf{X}}_1$ elements (before the **Ṕ**-slot). Seemingly simple 'fixes' to this problem, such as proposing **Ṕ Ṕ Ṕ** in place of **Ṕ** and **cl cl cl** in place of **cl** in the template gives rise to serious difficulties. For example, if multiple **Ṕ** or **cl** elements appear in some regular, well-defined order, then they must represent *distinct* positions in the template and thus *distinct* categories of syntactic object.

69. That is, if δέ *dé*, γάρ *gár*, and ἄν *án*, all presumably members of **Ṕ**, occur in a fixed order, we will need distinct (sub)categories, and the nice **Ṕ** in the above template will need to be replaced by $\acute{\mathbf{P}}_1$, $\acute{\mathbf{P}}_2$, $\acute{\mathbf{P}}_3$, …, $\acute{\mathbf{P}}_n$. Class membership definition will become increasingly stipulative if we do this, since now we will have to assign postpositives to multiple distinct ordered classes. Of course the same is true for the **cl** class if these elements also come in some regular order.

70. Wackernagel also needs more traditional (in the modern world) concepts like 'syntactic constituent', and, by implication, some precedure that creates such objects (i.e., a 'syntax'), apparently alongside his 'template' entity and 'slot-filling' procedure. This gives rise to the technical question of how the two systems interact, and why they interact in the way they do.

71. Finally, for Wackernagel, a given author could be guided by the 'strong' influence of the rule, or by a 'weakened' version of the rule. The rule is felt to have a life-span (strong at the beginning, weak later on). But the modern syntactician has to ask what the *grammar* of the speaker was at each stage and how 'strength' can be a property of the processes *within a grammar*.

2.3. Trouble in Paradise

2.3.1. 'schwach betonte Wörtchen'

72. Much of the machinery needed by Wackernagel, as sketched out at the end of the last section, violates some or all of the principles we've talked about in the preliminaries. In addition, the set of entities he needs seems like it may run afoul of our principles: while we might be willing to accept something like [+clitic] as a morphosyntactic feature (if we show that clitics can have a unique syntax—think of French object pronouns, e.g.), the idea that we are going to need [+cl_1], [+cl_2], [+cl_3], etc. (to account for ordering in the 'clitic chain') seems unlikely, and, together with [+postpositive_1], [+postpositive_2], [+postpositive_3], etc. begins to look like blatant stipulative data-matching. Recall that such 'data-matching' theories are the worst kind for a scientist, because they give the illusion of 'accounting for' the data, when in fact they are just recapitulating it, with no added insight (and no constraints on the range of possible systems humans can possess).

73. If we look again at Wackernagel's own definition of the law (cited above), several properties of that definition jump out at us immediately as potentially problematic. Let's take a look at a couple of those.

74. First, the *target* of the Law (i.e., the set of entities over which it is being claimed to hold) is given by Wackernagel as 'schwach betonte Wörtchen' ('weakly stressed little words'). Both of the concepts involved here, a weak degree of stress and the classification provided by the use of the *-chen* dimunitive, seem like they are phonological in nature. But we have argued at great length that the syntax *cannot see* the phonological properties of the entities over which it computes. When we made that claim, we were not doing so because it seemed like a fun thing to do: we were doing it because we have never seen a language in which all stressed words are moved to the front of the sentence (before all unstressed ones), or in which monosyllables must precede disyllables which must precede trisyllables and so on, or in which words that start with /b-/ must precede words that start with /t-/. We need to prevent the child/acquirer from analyzing the data to see if s/he is being confronted by such a system, or s/he will never successfully construct a grammar!

75. And recall that we cannot do this by explicit stipulative blocking of every relationship and every set of entities the syntax *cannot* pay attention to (there are an infinite number of these). Instead, we must state in positive terms what entities the syntax *can* operate on, and what operations it *can* perform. If we consider all of the properties that linguistic objects seem to possess, phonological, morphosyntactic, and semantic, it seems like the most sensible place to draw the line between what the syntax can 'see' and what it cannot would be between *morphosyntactic* features, which the syntax can access, and everything else (which it cannot). Indeed, this would be a sensible, perhaps the only sensible, way to define 'morphosyntactic feature'. If we define it this way, the question arises as to whether we are going to be forced to deal with an ugly world in which things we all agree are *not* morphosyntactic, are nevertheless to be called 'morphosyntactic' because of our theory. To the extent this is the case, we should be unhappy (unless we feel we have learned something truly new and exciting about the world).

76. So Wackernagel's definition of the target class is very problematic, given the architecture of the human mind (in the linguistic domain). Various 'solutions' to this problem have been proffered. First, one could declare that Wackernagel's Law is simply not a syntactic phenomenon at all, but rather a phonological one. As we will see, I think there is a kernel of truth in this idea — elements subjected to what we call 'WL' are all accessible to computation by the phonology in interesting ways, as, of course, are all elements that the syntax sends to the phonology. However, many proposed implementations of this idea are, in my view, deeply unsatisfying.

77. Second, one could argue that the syntax can see certain aspects of phonological structure (including degree of stress and 'Wörtchen' status, I guess), but not others, thus expanding the set of possible syntactic systems only along certain dimensions, but not as badly as one would do if one in general denied the modularity of the grammar.

78. The most difficult hypothesis to advocate would appear to be the one that seems to violate Occam's Razor (which I take very seriously): one could believe that the surface position of clitics is a function of (1) where they were placed in the syntactic structure, i.e., by the syntactic computation and (2) how that placement was subsequently affected by the phonological computation, if at all. That is, both domains could be relevant.

79. I say 'seems to violate Occam's Razor' because, of course, to the extent this theory relies on the interaction between a syntactic system that we need anyway, and a phonological system that we need anyway, and since we need the syntactic system to feed the phonological system (for postlexical phonology to work), it may come to us with no violation of Occam, since we would have added nothing additional to our explanatory arsenal.

80. There's an issue that we need to get clear about before moving on to consider these matters. It can be demonstrated with a famous pair of upper-class French liaison examples:

(3) a. un marchand de draps[z] anglais 'a merchant of English sheets'
b. un marchand de draps[∅] anglais 'an English sheet-merchant'

81. Or, for fun, here are some Sanskrit examples.

With preserved /s/:

(4) a. nutthấ ácyutaṃ sádasas pári svắt
you_pushed immovable-AccSg seat-AblSg PostP own-AblSg
'you pushed the immovable one from his own seat' (6.17.5d)
b. divás pári prathamáṃ jajñe agníḥ
heaven-AblSg PostP first born Agni-NomSg
'Agni was born first from heaven' (10.45.1a)

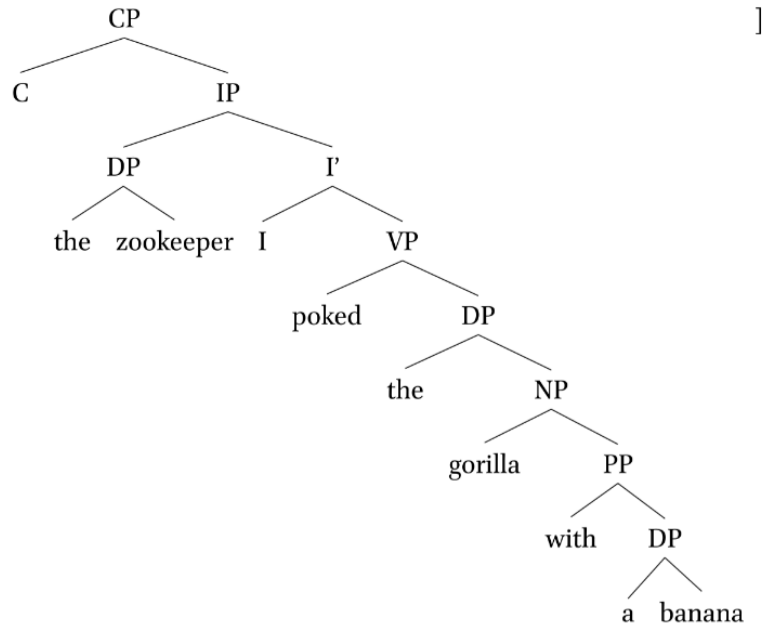
With visarga formation:

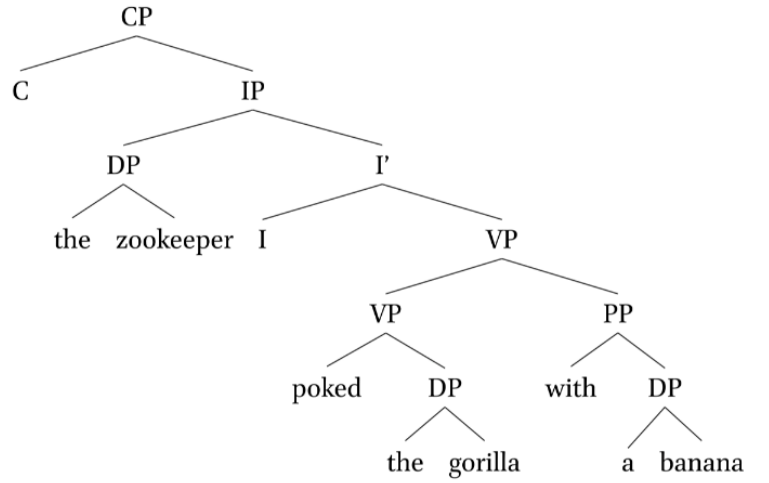
(5) a. sá kṣápaḥ pári ṣasvaje
he-NSg nights-APl PV embraces
'he embraces the nights around;' (8.41.3a)
b. mấ maghónaḥ pári khyatam
NEG generous-APl PV look
'do not overlook the generous one' (5.65.6c)

82. Let's start with the French examples. It looks like whether we are dealing with 'sheet merchants who are English' or 'merchants who deal in English sheets' is being *determined* by the phonology, in violation of the restriction on modular interaction I discussed above. It looks like a phonological event ([z]-preservation or [z]-deletion or whatever) has *semantic* effects.

83. This is a common point of confusion, so let's try to get clear about it before we move ahead. We need to distinguish between the information which is available *to us* when we hear or read a sentence, and the information that is available *to the grammar* when it computes over the elements in a sentence. A convenient way to think about this is to phrase it in terms of the *speaker's* situation relative to the string versus the *hearer's*.

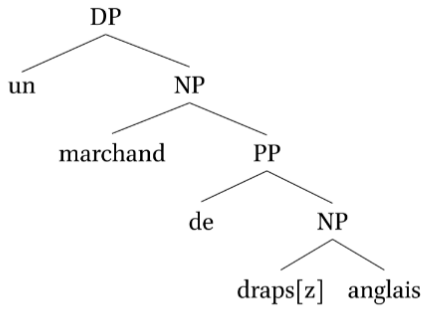
84. For example, I earlier cited the famous example of ambiguity: 'The zookeeper poked the gorilla with the banana.' I did not at that time point out strongly that the ambiguity is not a property of a *sentence*, but of an acoustic or orthographic form. The *sentence* of course is an object, built up by the syntactic computation, which encodes the grouping of the morphosyntactic objects in question. So, the *sentences* that this 'ambiguous' string encodes are two unambiguous ones (I won't worry about the internal structure of the Verb Phrase in the second example, which is not immediately relevant to our concerns, nor about the DP/NP issue, nor about much else).

a. 

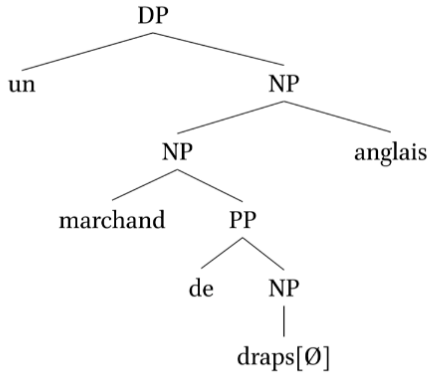
b. 

85. So the relevant part of the French examples would look like this:

a.



b.



86. And similarly, in the Sanskrit examples, we have something like [*divás pári*]$_{\text{PP}}$ as opposed to [[*maghónaḥ*] [*pári khyatam*]].

87. Okay, but haven't we just made our problem worse, in a way. We said the syntax can't compute over phonological features, and the phonology can't compute over syntactic features. So how does the fact that these trees differ help the phonology, which has to decide whether to tell the speaker to say [draz] or [dra]?

88. We have, thankfully, *not* made our problem worse, for a very simple (but as far as I can see virtually always overlooked) reason: the groupings indicated in the tree structure, created by the syntax in the course of the derivation, are not *morphosyntactic features*. The relationships created by the syntax are being created *for the semantic and phonological systems to interpret*!

89. The phonology computes over the phonological features provided by the elements in the syntactic representation in a manner consistent with the 'bundling' of those features which the tree represents. The semantics likewise, with respect to 'semantic features'. Indeed, the only function of the morphosyntactic features is to provide instructions to the syntax about how to 'bundle' the phonological and semantic features—the syntactic features themselves will be ignored at the end of the syntactic computation.

90. So when the phonology shows sensitivity to branchingness in a tree structure, it is not violating strict modularity, it is simply computing over bundled features, just as it must compute over the bundling of features in lexical items (every /g/ is a bundle of the features [+voice, -sonorant, etc.], in /kæt/ we have a bundled sequence of bundled features, in /ðə kæt/ we have a bundled set consisting of two distinct bundled sequences of bundled features, etc.
etc.).

2.3.2. Returning to the 'schwach betonte Wörtchen'' Problem

91. Let us now return to the first major issue confronting Wackernagel's own Wackernagel's Law: the problem of the *targets* of such a generalization. Here's part of Iliad 13.11-13. There are, in my view, 4 weakly stressed 'Wörtchen' (*gàr*, *te*, *te*, and *ep'*, underlined): 1 of them is in second position in its clause by a relatively simple, Wackernagel-inspired count.

(6)

καὶ	γὰρ	ὃ	θαυμάζων	ἧστο	πτόλεμόν	τε	μάχην	τε
kaì	gàr	hò	thaumázōn	hêsto	ptólemón	te	máchēn	te
and	for	he-NSg	marveling-NSg	sat	war-ASg	and	battle.-Sg	and

For he sat marvelling at the war and the battle,

ὑψοῦ	ἐπ'	ἀκροτάτης	κορυφῆς	Σάμου	ὑληέσσης
hypsoû	ep'	akrotátēs	koryphês	Sámou	hylēéssēs
high up	upon	highest.GEN	peak.GEN	Samos.GEN	wooded.GEN

high up on the topmost peak of wooded Samos,

Θρηϊκίης:
Thrēïkíēs:
Thracian.GEN
the Thracian one.

92. We do not get here, nor ever in Homer, or in any other Greek author, from such a thing a string like: *kaì gár te te ep'* …, where the weakly stressed clitics and postpositives appear in second position.

93. Wackernagel was not an idiot. It seems pretty clear that Wackernagel did not *intend* for the law to cover *all* weakly-stressed 'Wörtchen', in spite of what he wrote. Unstressed 'little words' such as ἐφ' *ep^h'* never show any kind of tendency to appear in second position, but their 'weakly stressed' status unambiguous, and you can't get much littler.

94. In addition, τε *te* 'and' always stays within the domain of the entity it is conjoining, and shows no 'clause-second' positioning tendencies (this is true of all the reflexes of IE **k^{w}e* 'and').

95. Even with these cases aside, there are countless 'exceptions' to Wackernagel Law amongst the elements he does expect the law to cover. And he knew that perfectly well, and cites many of them. He side-steps the empirical difficulties raised by the Homeric data we have just examined in part by assuming that Wackernagel's Law is a *probabilistic generalization* regarding the syntax of the archaic daughters he considers.

2.3.3. 'in sehr hohem Masse': WL as a Probabilistic Generalization}

96. We now turn to the second complicated issue which arises directly from Wackernagel's own characterization of 'Wackernagel's Law.' Wackernagel conceived of his law as a *probabilistic* generalization over the data. This is sometimes reflected in the way subsequent scholarship characterizes Wackernagel's Law, and sometimes it is not.

97. At any rate, as I said above, any survey of clitic distribution in archaic IE daughter languages, including that of Wackernagel himself, reveals exceptions to Wackernagel's Law. This entails that WL is either (1) blatantly false or (2) probabilistic in nature. Statistical generalizations have played a key role in many earlier approaches to IE syntax, indeed, in most. It is to the nature of such generalizations, and how useful they are to our scientific purposes, that I wish to turn now.

98. Let me remind you of the two purposes I have in sight in undertaking the study of historical syntax: (1) I want to understand the nature of syntactic change (and reconstruction), so that I can better understand the nature of the human genetic endowment in the linguistic domain (and thus the human mind generally); (2) I want to understand the meaning of the ancient texts which have been preserved for us. As I said at the outset, our theories should be structured so as to further our purposes.

99. Secondly, I noted at the beginning of these lectures that I want to try to use *science* to make progress on these goals, and that that commitment entailed certain requirements on the work we undertake. In my view, one of the core assumptions of the scientific enterprise is that, from the level of galactic drift to that of the Higgs boson, the causal structure of the world is intact. That is, the behavior of the objects which make of the world is driven by the forces which operate upon those objects, and we can formulate useful hypotheses about both the objects and the forces.

100. As I mentioned, the most popular approaches to IE syntax, particularly with respect to word order related phenomena, have been 'probabilistic' in nature. This includes both the general approach of Delbrück (e.g., 1900:38) and that of the SOV/SVO/VSO 'typological approaches of the 1970's. Delbrück's core idea was that calm, unemotional prose reflected the basic ('habitual') word order of a language, and that this basic order could be perturbed by excitement, drama, high emotion, and the like, which gave rise to 'occasional' word order patterns.

101. These generally coincide with later notions of 'markedness' or 'frequency-based assessment' of word orders, though the three categories ('habitual', 'unmarked', 'most frequent') can and sometimes do yield different actual results when applied to a corpus. In practice, the categorization has virtually always been based on corpus frequency data.

102. In identifying the statistically 'normal' attested word orders of a language there was a sense, in this approach, that something essential about the language had been captured —it was this 'essential nature' that was to be exploited for reconstruction purposes. As with many theoretic devices, we can treat this categorization system as a way of sorting the data: some sentences will represent 'habitual' (or 'unmarked' or 'more frequent') ordering relationships, some sentences will represent 'occasional' (or 'marked' or 'less frequent') patterns. One question we can ask fairly straightforwardly is this: does sorting the data in this way help us to achieve our goals?

103. It is very difficult for me to see how it possibly could. Assigning a label such as 'occasional', or 'marked', or 'rare' to an observed structure doesn't tell us anything particularly clear about what it means for the clause to be structured in that way, only about its corpus frequency (so it can't help with the 'text interpretation' goal).[7] And it can't help with the 'syntactic change/reconstruction' goal, because labelling a structure as 'occasional' or 'marked' or 'rare' doesn't tell you whether it represents a budding innovation or a precious archaism, only that it is has a certain frequency of occurrence.

104. Indeed, we can ask a more fundamental question: does it even make sense to ask about the frequency of certain word orders? Here is what I mean. There is a lot of literature about Indo-European syntax which counts the frequency of SVO sentences vs. the frequency of SOV sentences (for example) in a particular corpus. But why does no one ever talk about the relative frequency of SV (intransitive) sentences relative to SOV sentences? Intransitive sentences are more frequent, are they therefore 'habitual' or 'unmarked' and transitive ones 'occasional' or 'marked'? Why was there no discussion about whether PIE was 'SV', rather than 'SOV'?

105. Put another way, why does the SV:SOV contrast stand *outside* the scope of these statistical discussions, but, e.g., the OSV:SOV contrast stand within them? It seems that, implicitly, SOV and OSV are felt to be 'variants' of one another in a way in which intransitive SV is not a 'variant' of transitive SOV. Intuitively, keeping the transitivity alternation (SV:SOV) distinct from the SOV/OSV alternation seems to make sense. Speakers use transitive verbs when they want to express the kinds of predications which are expressed by transitive clauses and they use intransitive verbs when they want to express the kinds of predications which are expressed by intransitive clauses. Studying the relative frequency of the two types of clause (transitive/intransitive) in, e.g., the Vedic corpus, simply tells us how frequently the authors of the relevant texts wanted to express which kind of meaning. This is on the face of it no more 'syntax' than is a study of how frequently the authors wanted to express the meaning 'duck'.

[7] Of course, it could inspire us to subject a given structure to greater (or less) scrutiny, but it would be the results of that scrutiny—structural facts about the clause and its meaning in context—which would be helpful, not the mere fact of rare or frequent attestation. It is thus at best a heuristic, not the goal of scientific investigation, and not, of course, a property of a language, let alone its 'syntactic essence'.

106. But is it true that SOV/OSV are variants of one another in some way in which SV/SOV are not? Imagine that one way to get an OSV sentence in Vedic Prose is via focussing of the direct object. We talked about this briefly in the methodological section of these lectures, but let's recap what we might say about this. The analysis would seem to entail, for any model of syntax, that the speaker select (1) some element, call it the 'focus head', which establishes the location within the string of the focus position (in Vedic, near the front of the clause, on which more later) and (2) some element, call it a [+focus] feature, which serves as an indication to the syntax as to which element is to be focussed (so the speaker doesn't focus stuff unintentionally and at random).

107. It follows that a comparison of the statistical relationship between the OSV clauses generated in this way and SOV clauses which do not have focussed objects is just a count of how frequently the authors of our texts wanted to express the relevant meaning ('object focus')—i.e., it is conceptually no different than counting how often we find intransitive vs. transitive clauses. We are just counting 'non-object focus transitive clauses' vs. 'object-focus transitive clauses'. Two different kinds of clauses, that exist to express two different kinds of meanings. Put another way, comparing such OSV clauses to SOV clauses is actually (under the assumptions given) a comparison of 'O[+focus] FOCUS-HEAD S V' vs. 'S O[-focus] V'. And these are not simply different orderings of *the same* elements. The simple fact is that one could write an entire Neo-Rigveda using only OSV clauses *without changing the **syntax** of Vedic Sanskrit at all*. This is so for a very simple reason: the syntax doesn't tell you **how often** to express particular meanings, only **how** to express them, once you've decided you want to.

108. I am fond of citing in this connection a passage from Gilbert Sorrentino's novel Gold Fools (2001:249-50):

> At this rebuff, or congeries of rebuffs, did Hank 'splain, as clearly and humbly as possible, that a wash is a kind of gully down which, one fine day, ages past, a stream had done flowed among really old rocks? Was this inept description a sign of Hank's growing desperation? What were, say, two of the other signs? Was this putative wash a fur piece, further mebbe than Hank figgered? Did the frenzied chump urge his reluctant charges to put on some danged speed? Did Billee suddenly realize that a wash meant water? Or, at least, water hole? Again, was the mystery of linguistic legerdemain asserting itself? How, indeed, Nort mused, could "wash" even mean "water hole"? Was Billee flirting with the ineffable realm of the langue? Or the equally ineffable, and more expensive realm of the parole? Or had he, 'twixt sleep and wake, entered the world of "competence" and "performance"? Did Billee know what he was doing, or thinking, or was he the same know-nothing, Jesus-waving sod he'd always been?

109. Sorrentino's entire novel is written in the form of interrogative sentences, and yet it tells a story, with a plot, and a coherent sequence of events, easily discerned by any speaker of English (with awareness of the relevant dialects). Coherent syntax only requires syntactic well-formedness: usage rates depend on the speaker's communicative intent; the mapping between the two can be quite complex, as the Sorrentino passage clearly demonstrates.

110. We can contrast these 'counting' approaches with a type which seeks the explanations for word orders not in general markedness or frequency domains, but by trying to discover what *structural properties* are present in the strings, acknowledging that structural properties exist to express meanings. It seems pretty obvious that this kind of approach *will* definitely help with our broader goals.

111. Importantly, there is under this conception of things no such thing as a 'basic order' from which all other orders are derived: each order is derived from the elements which enter into the syntactic computation. A difference in order entails a difference in inputs (and a difference in interpretation). Equally importantly, it *does not* follow from what is said that there are not ordering relationships which get created in the course of the derivation only to be modified by subsequent operations. In our made-up and not justified thought experiment example above, the O[+focus] will have been inserted into the computation as a complement of the verb (so that it will get the interpretation which a direct object has) and will 'move to' (a metaphorical label for 'enter into a new relationship with') the FOCUS-HEAD later in the derivation (so that it will get the appropriate focus interpretation). Syntactic 'movement' is a label for the process which provides a single element with multiple such interpretations (since interpretations arise from the relationships created in the course of the derivation). We noted this earlier.

112. One might well counter at this point that the arguments against the 'typological' (SVO, SOV, OVS, etc.) approach to Indo-European word order may not be valid for some of our more challenging languages, such as Ancient Greek, Vedic Sanskrit, and perhaps a few others, but surely there is no denying that it is quite sensible to assert that Hittite is basically SOV, Germanic SOV with 'verb-second' effects, or Old Irish VSO. That is, while, as in many areas of Indo-European studies, serious empirical problems are still there to engage our attention, we do have some firm results from which admittedly provisional conclusions could be drawn. Am I not throwing the proverbial baby out with the likewise proverbial bathwater?

113. Unfortunately, I do not think that I am. The seemingly safe claims regarding the ordering of 'S', 'V' and 'O' in some archaic branches is, under careful scrutiny, also essentially unhelpful in our pursuit of useful knowledge regarding the syntax of Indo-European and its most archaic daughters. The basic problem concerns the interpretation of the significance of any given linearization (SOV vs. SVO vs. OVS, etc.).

114. A single observed order, e.g., SOV, may correspond to any of a number of distinct structural configurations. For example, there appear to be so-called 'verb final' languages whose verbs occupy some relatively high position in the tree (having moved to that position from their underlying one) and other 'verb final' languages whose verbs are *in situ* in final position within the verb phrase. If this contrast is real, stating that Hittite, e.g., is SOV fails to tell us *what kind* of SOV Hittite is—i.e., what is the expected *structural position* of V in that language in 'pragmatically neutral' clauses (if such a thing actually exists)?

115. Similarly, there may be many possible positions for 'S' in Hittite sentences, each of them showing S before VP-internal arguments, but each structurally distinct: the subject noun phrase may be focus-fronted to clause-initial position, it may be cliticized into the initial clitic chain (and thus occur before the V and its arguments), and, of course, it may be in normal 'subject' position. Each of these structures may show an "early S" surface linear order, but because they are distinct, and, one hopes, come with distinct interpretations, we need to establish, as best we can, just what structural configuration we are being confronted by in each case.

116. But if we have done that, if we have a theory of the structural positions of the elements making up the clause, what *additional useful information* does saying the sentence is 'SOV' provide us with? Linear order falls out from the structural configuration of elements. If Hittite shows SOV order with high frequency, it is because Hittite has a set of syntactic processes which place elements in the structural positions which give rise to such surface linear orders. If we understand the structures and the processes which give rise to an epiphenomenal order of elements such as SOV, what insight is provided by directly asserting that Hittite is characterized by this epiphenomenon (i.e., 'is' SOV)?

117. Put simply, linear order is the by-product of the input elements (the lexicon) and the syntactic system which computes over those elements: once we understand the inputs and the computational system, the outputs are predictable and thus tell us nothing additional. In telling us nothing additional, they cannot help with textual interpretation (those interpretations will follow from our understanding of the elements and the syntactic system), nor with our account of syntactic history. The correct historical grammar of the language is not a question of how we get it to have a set of superficial, epiphenomenal properties, but how we get it to have the set of morpho-syntactic objects and processes that are actually responsible for these epiphenomena.

118. Finally, there is a fundamental challenge confronting any such statistical approaches to syntax. Probabilistic systems do not explain any *individual* piece of data, because, if they could, they would not be probabilistic. Such systems only make claims regarding the statistical behavior of a *set* of data. Think of a simple probabilistic generalization: for the past 25 years, I have on average come to Edinburgh once every five years (i.e., in 20% of the years). What does this tell us about my behavior with respect to visiting Edinburgh? Am I in Edinburgh this year because of this statistical generalization? Of course not, I'm here because I was kindly invited and was able to accept. In years that I am not in Edinburgh, am I not in Edinburgh because of the statistical generalization that I am only here in 20% of the years? Of course not; there are lots of the reasons why I am not here in any particular year, but *none* of those reasons are provided by the statistical analysis of my behavior!

119. The statistical generalizations regarding my behavior, or the behavior of subjects, objects and verbs in archaic IE language clauses, are a *result* of the interaction of the actual causal processes that *do interact* to determine observed word order patterns. Because statistical models do not make claims regarding individual pieces of data, they are inherently discouraging of further research into factors that might be responsible for a given example, and thus discourage the research which would show that the assumption of an underlying probabilistic dynamic is false. And, of course, as already pointed out above, they do not give us a meaning for our texts, nor do they help our understanding of the data in a way that helps us to model reconstruction or change.

2.3.4. 'welches immer ihre syntaktische Beziehung sei'

120. This quote from Wackernagel's definition of his law expresses the belief that the position of 'weakly stressed little words' in second position takes place in complete independence from the actual syntactic relationships the clitic may have with other elements of its clause. But recall our earlier example in which *te* and *ep'* **failed** to move to second position. They instead surface adjacent to their arguments. Isn't this trying to tell us that their syntactic relations *are* relevant?

121. So, why do *some* 'weakly stressed little words' show up in second position of their clause and others not? Can be we less stipulative about what kinds of 'weakly stressed little words' show up in 'second position'?

122. Obviously, one cannot define these non-WL elements *prosodically*—that's in some sense the whole point of the above example. These elements have the *correct* prosody; they just aren't subject to Wackernagel's Law (at the clausal level). Since we really only have two domains which could even in principle be responsible for the placement of elements in the clause—the syntax and the phonology—if the definition of these elements does not follow from their phonology, it better be a function of their syntax. But if some weakly stressed little words go to second position, and some do not, and the contrast is based on the *syntactic properties* of the elements, then Wackernagel's 'welches immer ihre syntaktische Beziehung sei' cannot be correct. Nor, by the way, can any purely prosodic approach to the distribution of WL elements (e.g., Keydana 2011).

123. Note that this observation could provide a partial explanation for why Wackernagel's Law is seen by many (including Wackernagel) as *probabilistic* in nature. Since the Law is defined, as a probabilistic generalization, only over a set of 'weakly stressed' elements, it ignores the *syntactic structure* within which these elements are found. But if syntactic structure imposes conditions on extraction, and moving up towards the start of the clause (as would seem to be required by WL) *is* extraction, then ignoring syntax can only lead to an approximation of the phenomenon. Since one's analysis is only approximating the correct analysis, it is necessarily probabilistic in nature. But what we want to know is what the *actual* correct analysis is.

124. In the final section of these lectures, I'll try to get us at least a little closer to 'historical linguistics heaven' by focussing on what we **can** do with what we've posited, rather than complaining about the shortcomings of the work of others. This is a little unnatural for me, so bear with me!

3. Tercia pars, que dicitur Paradisus.

3.1. Okay, smartypants, if there's no WL, why, whenever I read an archaic IE language, are there so many clitics and postpositives in second position?!

125. There is no rule that says that 'weakly stressed little words' must, or even tend to, in any interesting sense, go in clause second position regardless of what syntactic relationships they have in their clause. Conjunctions like reflexes of IE **k^{w}e* do so only under specific and well-understood conditions (likewise disjunctions like Vedic *vā* or Latin *=u*), and the Attic Greek definite article and proclitic prepositions likewise resist any such pressure. This is one clear sense in which there is no such thing as Wackernagel's Law at least in the form given by Wackernagel himself.

126. But there is another sense in which there is no such thing, and for that, we need (finally) to turn to what factors we think *are* relevant to our understanding of the distribution of clitics and postpositives in archaic Indo-European languages. What I will argue is that once we do understand these factors, none of them will turn out to be identifiable as 'Wackernagel's Law' in the sense that (1) all of the relevant factors are in play in the behavior of elements which are *not* subject to Wackernagel's Law (in Wackernagel's sense, i.e., regarding 'clause-second' position) and (2) none of the relevant factors plays a decisive role in placing an element in 'second position' in all of the cases of seeming compliance with Wackernagel's Law.

127. The end of the last section gave us the critical hint as to how we might approach these issues. Syntactic structure plays a critical role, not only in placing the elements in their ultimate position, but also in serving as a limiting force on the ability of a clitic or postpositive to end up high in the clause. On the other hand, the fact that 'second position' is defined as 'after the first phonological word' clearly entails that the phonological computation also plays a crucial role in giving rise to second position phenomena.

128. Therefore, the observed data are to be explained as arising due to the interaction of the syntactic system and the phonological system. Recall that this interaction does not arise in a parallel or interculcated computation, with the two kinds of structure being built up simultaneously, or in an interactive manner. Instead, under the assumptions I have made, the syntax *first* constructs a representation, then, this representation is passed on the phonology for phonological computation. So, what are the relevant syntactic and phonological processes that we need to think about in developing an analysis of Wackernagel's Law-type phenomena?

129. Let's recall some principles we've adopted regarding syntax, because they are about to matter to us a good deal. First, there is only one syntactic operation: combine X and Y to make Y an argument of X (i.e., to put Y in the scope of X). So, it will combine a preposition and a noun to make the noun (or noun phrase, of course) an argument of the preposition. I will combine a verb and a noun to make the noun an argument of the verb. It will combine 'and' and 'Mary' to make 'Mary' the argument of 'and'. And so on. The most restrictive stance we can take on what syntax does is that it does *only* this. Secondly, the syntax is universal and invariant (only the mental lexicon differs from linguistic system to linguistic system).

130. It follows from these two assumptions (either or both of which could be wrong, but which, together, give us the most restrictive theory of syntax possible, I believe) that there can be no syntactic 'Wackernagel's Law' just as there is no syntactic 'relative clause formation' or syntactic 'question formation'. The syntax can compute over morphosyntactic objects by doing the only thing it does, concatenation, over and over, and in the end it can have made something which you and I may call a 'relative clause' or a 'question', but it did not do that by means of a special operation ('relative clause/question formation'), because there are no special operations.

131. Moreover, there would be no way for the syntax to put an element in a position that it could guarantee would *end up* being 'second' in the clause. When the syntax puts an element in the tree, which it is building from the bottom up via concatenation events, it simply doesn't know yet what is going to happen in the rest of the computation. That would require look-ahead, i.e., that before the syntax builds the tree, it needs to build the tree to see where things are going to end up, but you can't build the tree before you build the tree! So if the syntax puts the dative clitic *me* at the top of the tree at some point (which it is going to need to do to get it up near initial position), it can't know *at that point* how many more things are going to get put in above it—i.e., it can't know whether this is going to *end up* second position or not.

132. In the area of phonology, there is only one relevant operation that we will need to worry about in characterizing the data under investigation. Let me present an argument that the operation is needed, and a little bit about what its form is.

133. One thing that everyone has always agreed on is that there are sets of elements in archaic Indo-European languages which may not appear in clause-initial position (the clitics and postpositives). Such elements appear to share a single *prosodic* property,[8] which Wackernagel called being 'weakly stressed'.[9]

134. This prosodic fact is good for us, in a way, because we have already asserted that the syntax doesn't know how to put things in 'second position', so we are going to want the phonology to be involved in some way. But if the things that were going into second position differed from those that were not only in their *morphosyntactic* features (and not in their phonological ones), the phonology could not identify these elements (since it can't see morphosyntactic features) and it could not play a role in their placement.

135. One way to figure out what the syntax of a given element should be like is to take a look at semantically parallel elements. For the 'functional particles' which undergo Wackernagel's Law we often have directly parallel (and presumably slightly more emphatic, thus not prosodically deficient) elements. Beside the enclitic conjunction *ca* in Sanskrit, *-que* in Latin, and τε *te* in Greek we have the non-enclitic conjunctions *utá*, *et*, and καί *kaí*, respectively. I won't bore you with examples, but suffice it to say that if you want to say 'and Marcus' in Latin (or similar things in Greek, or Sanskrit) you have two choices: *et Marcus* or *Marcus=que*.

136. And for the enclitic subordinators Sanskrit *hí* and Greek γάρ *gár* (both 'because, since'), we have tonic subordinators of similar meaning: *yád* ('in that, because, since') and ἐπεί *epeí* 'because, since', respectively.

[8] There is a risk of circularity here—since we sometimes appear to be determining the prosody based on the distribution, and then using the distribution to justify the prosody we have assumed, so we must be cautious. More on this later.

[9] We must also keep logically distinct two very different claims: (1) Wackernagel's Law applies to (all) weakly-stressed little words vs. (2) only weakly-stressed little words can be subject to Wackernagel's Law. The first says that if you are weakly stressed you must follow WL, the second says if you follow WL you must be weakly stressed.

(7) a. imé cit táva manyáve
these-NDu even$_{cl}$ your fury-DSg

vépete bhiyásā mahī́
tremble fear-ISg great-NDu

yád indra vajrin ójasā
since Indra-VSg having-the-vajra-VSg might-ISg

vr̥trám marútvām̐ ávadhīr
Vrtra-ASg accompanied-by-the-Maruts-NSg you slew
Even these two great (world-halves) tremble with fear before your fury, since you, O vajra-bearing Indra, slew Vrtra with (your) might, accompanied by the Maruts.
RV 1.80.11a-d

b. sugó hí vo aryaman mitra pánthā
easy-going-NSg **since** your$_{cl}$ Aryaman-VSg Mitra-VSg path-NSg

anr̥kṣaró varuṇa sādhúr ásti
harmless-to-men-NSg Varuṇa-VSg straight-NSg is

ténādityā ádhi vocatā no
therefore-Ādityas-VPl PV speak for-us$_{cl}$
Since your path is easy, harmless to men (and) straight, o Aryaman, Mitra, and Varuṇa, therefore speak in our defense, Ādityas. RV 2.27.6a-c

(8) a. ἐπεί μ’ ἀφῆκε καὶ κατηδέσθη κτανεῖν,
epeí m’ aphêke kaì katēidésthē ktaneîn,
since me$_{cl}$ it released and was ashamed to kill
Since it (the city) released me and was ashamed to kill (me)… Hcld. 1027

b. Δέσποιν’, ἐγὼ γὰρ ἔρχομαι κατὰ χθονός,
Déspoin’, egṑ gàr érchomai katà chthonós,
lady-VSg I-NSg **since** I am going beneath earth-GSg
‘My Lady, since I am going beneath the earth,… Alc. 163

137. It is instantly noticable that in the case of such pairs, the relationship between their surface position can be captured quite simply: the enclitic version appears exactly one phonological word to the right of where the non-enclitic version would be expected to appear. So, if tonic *yád* had been used instead of *hí* in RV 2.27.6a, we would have expected *yát sugó vo…*. And if ἐπεί had been used in Alc. 163, we would have expected ἐπεί ἐγὼ…

138. So, what we know about these kinds of elements is this: (1) their surface position is at least in part the responsibility of the phonology, (2) they cannot appear clause-initially, and (3) they surface one prosodic word to the right of their expected syntactic position. These three considerations lead to a relatively straightforward theory:

> When an enclitic element (which must lean to its left) sits at the left edge of its domain (and thus has nothing to its left to lean on in that domain), the phonology ‘repairs’ this structure by a *minimal* rightward movement of that element.

139. By minimal, I mean 'one prosodic word'. The 'within their domain' codicil will be explained as we move ahead.

140. What does this mean? Well, it means that the structures we mentioned above did not have the same linear order at the end of the *syntax* as they did at the end of the phonology. At the end of the syntax, we would have had strings as below (I've used = to indicate an element must lean to its left).

(9) a. et Marcus …
b. =que Marcus … (→ Marcus=que …)
c. utá vr̥trám …
d. =ca vr̥trám … (→ vr̥tráṃ=ca …)
e. kaì paîs…
f. =te paîs … (→ paîs=te …)

(10) a. yát sugáḥ …
b. =hí sugáḥ … (→ sugó=hí …)
c. epeì egṑ …
d. =gár egṑ… (→ egṑ=gár …)

141. You can see how advantageous this is: we now get very similar semantics from very similar syntactic structure, with no additional machinery in the mapping from syntax to semantics. We require additional machinery in the phonology, but since the differences that seem to trigger the re-ordering are prosodic in nature, we have placed the machinery in the domain that seems most appropriate (given modularity).

142. The next question we must confront is whether we can call this phonological operation (which I will, following Halpern's 1990 Stanford dissertation, call the 'prosodic flip' or 'prosodic inversion', though neither term is great, since they don't make it clear that *only* the prosodically deficient element is being repositioned) 'Wackernagel's Law'. If we can, then the Law exists, and is a phonological (not syntactic) operation. But I will argue that we cannot, if we are to use the term 'Wackernagel's Law' to refer to the data that Wackernagel himself presented in support of the Law.

3.2. Syntax, Prosody, and Wackernagel's Law, again

143. If we believe what has been said up to this point, we have two processes which conspire to trigger apparent Wackernagel's Law phenomena: (1) syntax (universal and invariant) and (2) the 'prosodic flip', which shifts an enclitic which needs a host on its left *one phonological word* to the right if it would otherwise be unhosted.

144. We immediately see why some of the 'weakly stressed little words' which serve as exceptions to Wackernagel's Law, do not show up in second position. Prepositions and forms of the Attic Greek definite article are *proclitic*, not *enclitic*—i.e., they lean to the *right*, so they will never undergo 'prosodic inversion' at the left edge of their domain, the way that enclitics elements do. And of course they are fine in clause-initial position.

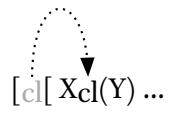
145. In addition, there is no syntactic operation in any human language known to us at this point which moves a conjunction (like 'and', 'und', etc.) up towards the head of the clause, away from the elements it is coordinating. There is thus no mechanism which would allow Sanskrit *ca* or Attic Greek τε to move up near the 'Wackernagel' position. It thus does not appear in clause-second position except accidentally.

146. Of course, it is still enclitic, and must be hosted on its left. For this reason, when the phonology gets a representation like $[\tau\varepsilon\ \pi\alpha\hat{\iota}\varsigma]_{NP}$, in which the τε in conjoining the noun παῖς to something, and thus takes NP (or DP, if you prefer) as its domain, the τε will have nothing to lean on within its domain to its left. It must therefore undergo the non-syntactic 'prosodic inversion' with παῖς, producing the attested order παῖς τε.

147. Note that, because Wackernagel expressly does not include τε in the set of forms his law covers (because the law says the elements targetted by the Law move to *clause-second* position, and τε does not), the 'prosodic inversion', which is responsible for the post-syntactic repositioning of *both* Wackernagel elements such as γάρ and non-Wackernagel elements such as τε cannot be Wackernagel's Law.

148. If the positioning of the WL clitics is a function of the interaction of this prosodic repositioning, which can't be Wackernagel's Law, and of 'normal' syntax, which can't be Wackernagel's Law, then there is no work left for Wackernagel's Law to do—there is nothing left for Wackernagel's Law to be.

149. There is another interesting implication of this approach to the phenomena traditionally called 'Wackernagel's Law': if the 'prosodic inversion' is the only displacement from the syntactic positioning, and it takes place only when the clitic does not have a prosodically suitable host to its left within its domain, then there are only *two* possible placements, for any given clitic: (1) X **cl** (Y)..., where the clitic is *properly positioned* by the syntax and X is *syntactically* to its left, within its domain (and thus a suitable host), and (2)



> where the clitic occupied a position at the end of the syntactic computation which did not allow it to 'lean' to the left, thus triggering 'prosodic inversion' (a phonological repositioning)

150. That is, every clitic under this restrictive hypothesis is either (1) in the position the syntax placed it in, like any other element in the clause, or (2) exactly one phonological word to the *right* of the position the syntax placed it in. No other positionings are expected. The key concepts involved in this theory are thus two: the expected syntactic position of the clitic and (2) hosts to the left of the clitic construed as being 'within its domain', which allows the clitic to lean on such hosts.

151. Let's walk through some clitic phenomena through this lens, and at the same time I will try to show you an interesting example of the role of UG in the *diachronic* development of clitics.

3.3. Clitics and Syntactic Change

152. Being able to reconstruct involves developing a sense for certain aspects of change: directionality limitations (X>Y but *Y>X), triviality (how likely is parallel independent development?) and, of course, possibility (can X>Y?).

153. Here I will propose a general, schematic account of the genesis of second position clitics and, more importantly, note that that account would *in principle* allow for the coming into being of a certain type of clitic which does not, however, appear to be attested.

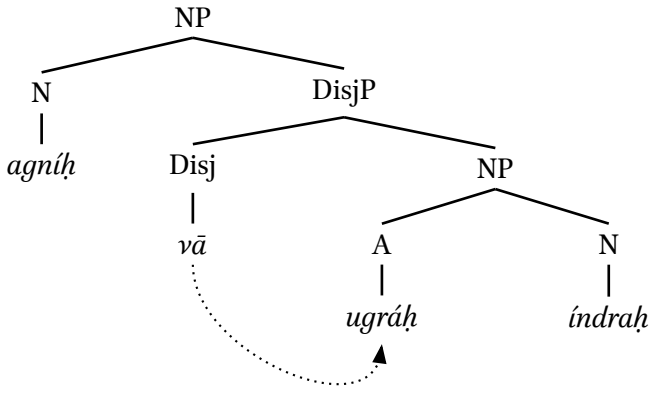
154. The non-attestation of that clitic cannot be attributed to the impossibility of an appropriate diachronic trajectory *except* by the assumption of a general grammatical constraint (a UG constraint, if you will) which precludes it.

155. This reveals that a key mechanism in making progress on constraining change, and thus enabling reconstruction, is to understand as fully as possible the two key elements of acquisition: ambiguous input data and a robust, constraining innate computational system (UG).

156. Clitics are prosodically deficient and require a non-deficient prosodic host. A given clitic lexeme will normally have to specify in its phonological representation whether it is to be hosted on the left or on the right. If, when the representation gets to the phonology (and the clitic is thus in its final 'syntactic' position), no appropriate host is available, the clitic may move (minimally) *in the phonology* to a position in which it can be hosted (a Calabrese-style phonological 'repair'). Distance in such cases is computed linearly (i.e., hierarchical structure is irrelevant). This is the so-called 'prosodic flip' of Halpern (1992).

(11) a. agnír ugró vā índraḥ
Agni mighty or_{cl} Indra
Agni or mighty Indra constructed ex.

b.



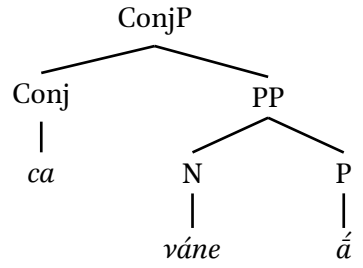
157. Sanskrit *vā* 'or' (like Sanskrit *ca* 'and', Latin *-que* 'and', etc.) requires a prosodic host on its left (thus it is never attested phrase-initially, for example). Its cliticization domain is the XP it is disjoining. In the tree above, *vā* sits at the left edge of that domain, and is thus not properly hosted. The *minimal* prosodic 'adjustment' which will resolve this ill-formedness, given that *vā* must remain in its prosodic domain, is a rightward shift over *ugráḥ*.

158. Note that *vā* occupies a surface position which no element of its syntactic type could *ever* occupy for purely syntactic reasons – there is no position within the NP for disjunction operators linking two NP's. Further evidence that the adjustment is prosodically-driven comes from examples such as RV 9.89.1cd, which involves the coordination of *mātúr upásthe* 'in the lap of the mother' and *vána á́* 'inside the wood':

(12) a. sahásradhāro asadan ny àsme
1000-flowing-NSg sat down us-LPl

mātúr upásthe vána á́ ca sómaḥ
mother-GSg lap-LSg wood-LSg P and$_{cl}$ soma-NSg
thousand-flowing Soma sat down by us in the lap of the mother and inside the wood
RV 9.89.1c

b.



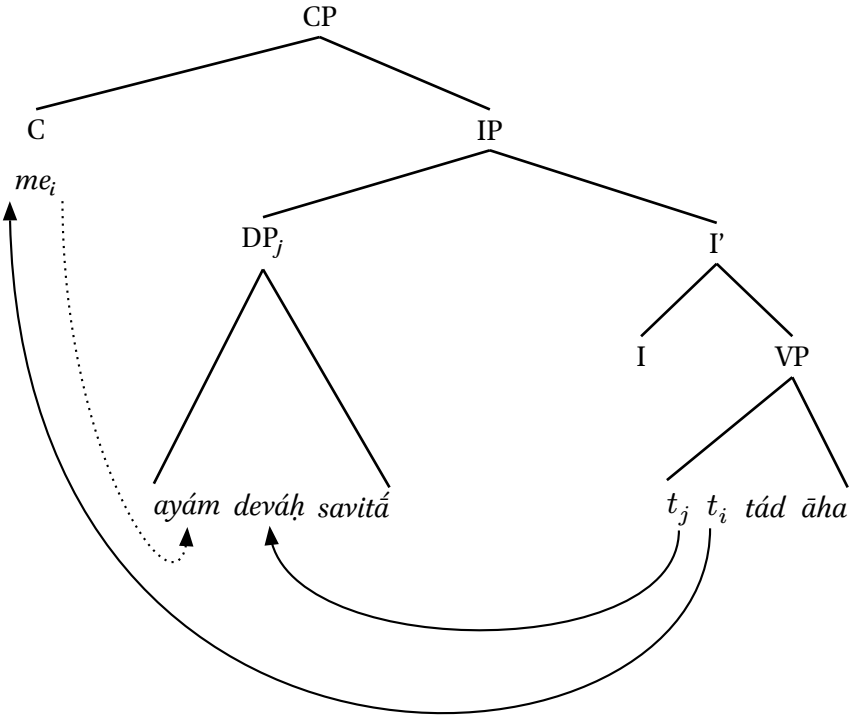
159. Postpositions which govern non-branching N's in Sanskrit form a *phonological word* with that N. We can see this in the resyllabification (and thus the failure to undergo debuccalization to [h]) in *divás pári* 'from heaven.' The preservation of the *s* of *divás* finds a straightforward explanation if it is the case that this *s* is *internal to the phonological word* after phonological words are constructed and does not, therefore, meet the conditions ('word final') for the debuccalization.

160. Given this, *vána á́* should also represent a phonological word. The first possible host for *ca* in the tree above is thus not *váne* but rather the newly-created phonological word *vána á́*. If *ca* placement were *syntactic* in nature, this would make no sense.

161. Since the clitic pronominals which move to C syntactically are also prosodically deficient, they must be hosted. If they are not properly hosted when they emerge from the syntax, they will undergo prosodic adjustment in the phonological component and thus be doubly dislocated — once through the syntactic operation which adjoined them to C and once through the phonological process which 'flips' them over the first available host.

(13) ayám me deváḥ savitá́ tád āha
this-NSg me$_{cl}$ god-NSg Savitr̥-NSg that-ASg said
'this god Savitr̥ said that to me' RV 10.27.18c

(14)



162. Note that the 'prosodic flip' will place the clitic in 'second position' thus giving rise to apparent satisfaction of Wackernagel's Law, though now for a completely different reason than we saw above where the clitic was properly hosted at the end of the syntactic derivation.

163. The term 'special clitic' (Zwicky) obscures crucial structural and computational differences between at least three distinct types of element

- clitics which have only undergone syntactic movement (and are thus 'displaced' from the position of non-clitic arguments) [e.g., some pronominal clitics discussed above]
- clitics which have not undergone displacement from their 'expected' position in the course of the syntactic computation but which, at the end of that computation, are not properly hosted (in the prosodic sense) and which, therefore, undergo 'prosodic adjustment' during the course of the phonological computation [e.g., *vā* 'or' and *ca* 'and' above]
- clitics which have undergone syntactic movement (and are thus 'displaced' from the position of non-clitic arguments) which *additionally*, not being properly hosted (prosodically speaking) at the end of the syntactic computation, undergo movement in the course of the phonological computation [i.e., 'doubly-displaced' pronominal clitics discussed above]

164. Terminology which mixes unlike entities, computational procedures, etc. obfuscates rather than clarifies. Therefore, we will refer to

- clitics which undergo syntactic displacement as 'syntactic clitics', bearing in mind that like all clitics, syntactic clitics must be hosted and thus may undergo further displacement in PF,

- clitics which occupy their 'expected' syntactic position at the end of the syntactic computation as 'simple clitics,' bearing in mind that because of prosodic displacement rules in PF such clitics may no longer occupy their expected syntactic position in the PF-output,
- clitics of either of the above types which have undergone prosodic displacement in PF as 'prosodically displaced,' thus 'prosodically displaced syntactic clitics' and 'prosodically displaced simple clitics'.

165. Examples include the following.

- **non-prosodically-displaced simple clitic:** English *'em*
- **non-prosodically-displaced syntactic clitic:** Sanskrit *me* 'me' without 'prosodic flip'
- **prosodically-displaced simple clitic:** Sanskrit *ca* 'and' or *vā* 'or'
- **prosodically-displaced syntactic clitic:** Sanskrit *me* with 'prosodic flip'

166. Let us turn next to the question what can, and what cannot, be a clitic? This will help us understand the diachrony of this category of objects. Given what we as linguists know it present the facts seem to be as given below.

- **Can:** pronominals, auxiliaries, modals (including 'evidentials'), 'adverbial particles' (including 'interrogative' particles), articles (D), conjunctions, disjunctions, emphatic 'particles'
- **Can't:** 'open class' lexemes (N, V, A, ...)

167. Okay, so why can('t) things that can('t) be clitics (not) be clitics?

168. To approach this question, we need to ask what factors go into the construction of prosody in the phonology. Here are some of my assumptions:

- there is a set of algorithms which construct prosodic domains on the basis of the syntactic output representation which is sent to the phonology
- these algorithms are sensitive to the groupings of elements established during the construction of the sytnactic tree (i.e., indirectly, to syntactic relations)
- these algorithms are sensitive to the lexical ('open' class) vs. functional status ('closed' class) of a given head (Selkirk 1984)
- these algorithms are sensitive to the prosodic properties of the lexical items in the tree (thus phonological 'movement' of unhosted clitics)

169. So, in Sanskrit we predict distinct degrees of sentential stress (itself a function of the domains created by the prosodic algorithms) for the indefinite instrumental singular clitic *tvena* (which is prosodically deficient *and* a 'functional' head), the definite instrumental singular determiner *téna* (which is tonic, but still a 'functional' head), and the neuter noun *téjaḥ* 'brilliance' (which is neither prosodically deficient nor a 'functional' head).

170. Further: as doubtless already mentioned, we need to distinguish between two different kinds of information which much be stored regarding lexical items:

- **lexicon 1:** encyclopedic (this is what tells us that 'duck' means 'duck' and not 'desk')

- **lexicon 2:** grammatical (this contains only information relevant to grammatical computation)

171. There are some compelling arguments that these are distinct:

    - 'Encyclopedic' lexicon learning violates the subset principle (i.e., allows acquirers to narrow originally overbroad hypotheses), 'grammatical' learning appears not to;
    - 'Encyclopedic' lexicon learning appears to lead, to at least some degree, to 'explicit' knowledge, whereas 'grammatical' learning does not.
    - 'Encyclopedic' lexicon learning appears to continue throughout one's lifetime, whereas 'grammatical' learning (of your native grammar) in my opinion does not.
    - The information in the 'encyclopedic' lexicon plays no role in grammatical computation.

172. So, given these arguments, we can ask: why can't 'duck' be a clitic?

173. There are actually two questions lurking here: (1) why can't 'duck' be a clitic in a grammatical system in which 'banana' is not? and (2) why can't 'duck' (or any N) be a clitic in any grammatical system? We can answer this question by asking ourselves this: 'where do clitics come from?'

174. First, recall that there are two principled explanations for why X isn't found in any human language:

    - UG does not allow X
    - there is no way for X to come into being diachronically (the 'diachronic' filter, recall the back vowel height issue)

175. Returning to our questions above in light of the forgoing we get:

    - Is a lexical item like *duck* [+N,+clitic] (or whatever) *computationally intractable* – i.e., out because of conditions imposed by UG?
    - Or, is a lexical item like *duck* [+N,+clitic] out because of the 'diachronic filter' (i.e., no evidence can, given existing conditions, give rise to a grammar that has this property)

176. Let us assume the following: to be reanalyzed as a 'simple' clitic, which is a 'prosodically deficient element', a lexical item must occupy a prosodic 'trough' in the phonetic output. The 'encyclopedic lexicon' (i.e., meaning) differences between 'duck' and 'banana' are not visible to the grammatical computation (no language has 'duck' fronting or 'banana' movement). Those aspects of the 'grammatical lexicon' representations of 'duck' and 'desk' which are used by the prosodic domain-constructing algorithms are the same.

177. The prosodic domain-constructing algorithm results in some elements — allegedly a function of the elements status as a 'lexical' vs. a 'functional' head (and computationally independent of clitichood) — receiving 'lower' degrees of sentential stress, others 'higher' degrees of sentential stress. (But 'duck' and 'desk' fall on the same side of this divide.)

178. Nouns (which come into the prosodic computation as 'lexical' heads) will not regularly occupy positions in the constructed prosodic domains in which they receive particularlyl 'low' sentential stress.

179. By contrast, tonic *functional* heads will regularly occupy 'troughs' in the sentence prosody, making their lexical tonicity harder to perceive than that of an open class lexical item, and more likely (therefore) to be reinterpreted as atonicity — i.e., prosodic deficiency, presumably as 'simple' clitics (because they are not *yet* clitics, so their positioning pre-reanalysis must be syntactically licit). And nouns, given their failure to occur in 'low sentential stress contexts, will not be easily reanalyzed as prosodically deficient (i.e., as 'simple' clitics). Like the 'front' vs. 'back' height distinctions, this is a function of (1) initial conditions (i.e., the fact that there are not N clitics now) and (2) diachronic possibilities. It cannot be used to deduce computational (or representational) properties of UG involving clitics.

180. The trajectory we are moving toward looks like this: tonic functional element → simple clitic → special clitic. But how do we get 'special' clitic step?

181. Imagine the ancestor of a simple sentence like *tvám mā kāmáyāse* 'you should love me' (RV 10.124.5b with its *ca* and its vocative removed). We have an emphatic subject pronominal in a pro-drop language, occupying initial position in its clause, following by an unstressed accusative object. In this ancestral structure, *mā* occupies the same position a noun like *índram* 'Indra-AccSg' would—it is a 'simple clitic' like English *'em*.

182. But, it is perfectly consistent with the analysis it actually has in the Sanskrit of the Rigveda (otherwise, the string couldn't occur in Vedic, which it can!): a 'special' clitic which has moved to the C-domain (like all 'special' pronominal clitics do).

183. In fact, such reanalysis may have been blocked for certain syntactic structures by syntactic considerations (in spite of Wackernagel's claims that syntax is irrelevant). In the ancestor of something like: *práti **mā** rócamānā marútaḥ* 'shining (their) light on me, the Maruts…', the clitic may *not* have been ambiguous. If there was a syntactic constraint against 'special clitic movement' out of participial clauses (entirely plausible, if not certain), then the listener could only have treated *this* instance of *mā* as an 'in situ', i.e., 'simple', clitic.

184. That allows us to make some sense of the apparent 'exception' to Wackernagel's Law in RV 1.165.12ab: *evéd eté práti mā rócamānā / ánediyaḥ śráva éṣo dádhānāḥ* 'thus indeed, shining (their) light on me, are these blameless ones establishing fame and nourishment'. The reason *mā* is 'violating' Wackernagel's Law here may be because *mā*'s in this syntactic position had to continue to be analyzed as 'simple clitics' and thus couldn't be thought to have 'gotten out' of their constituent of origin (and thus still now cannot).

185. Finally, let's turn to a case which has *never* been re-analyzed as involving 'special clitic' status: Sanskrit *ca* 'and'. Our two possible explanations for why Sanskrit *ca* cannot be a 'syntactic' clitic would be:

- a grammar in which *ca* 'and' undergoes syntactic movement to the C domain (like the 'syntactic' pronominal clitics) would be computationally intractable (i.e., is UG blocked)
- a grammar in which *ca* 'and' undergoes syntactic movement to the C domain cannot come into being diachronically

186. But examine the sentence in the next example.

(15) a. …yé ca rātíṃ gr̥ṇánti
…who-NPl and$_{cl}$ gift-ASg sing
…and who sing (you) a gift 4.34.10d

187. In this example *ca* is prosodically deficient and it occupies a position trivially reanalyzable as 'special clitic' movement to the C domain. In fact, the vast majority of clause-connecting *ca*'s could be taken as having been moved to the C domain. Diachrony thus does not preclude the reanalysis of conjunctive/disjunctive clitics as 'syntactic' clitics—the position of *ca* in the above string is in principle *just as ambiguous* as that of *mā* in the hypothetical example I constructed above. Why didn't listener decide that (at least clause-conjoining) *ca* was a 'special clitic'?

188. Diachronic reanalysis doesn't *a priori* preclude such a development, but **UG does.** There are no mechanism of syntactic computation which would allow conjunctions to move out of their domain to adjoin higher functional heads made available to humans. The absence of such mechanisms constrains the hypothesis space of the learner: they can only entertain theories of clause structure which are consistent with how human languages actually work.

189. The story of *mā* is designed to tell us that it is not appropriate even for our living languages to think in terms of a clitic element being *either* a 'simple' clitic or a 'special' clitic. Instead, we have a clitic element, whose behavior is determined by a set of conditions on movement. It moves up into the C domain when it has to and it stays within its constituent of origin when it cannot or need not so move.

190. When we add in the failure of *ca* (or any of the many reflexes of PIE *k^{w}e) to develop into a 'special clitic'—i.e., to be reanalyzed as having undergone syntactic movement—we see that two critical factors determine the nature of the constructed grammatical system: the set of input structures presented to the acquirer for analysis (the data) and the set of innate principles that regulate the construction of human grammars (UG).

191. This is the direct linkage between historical linguistics and theoretical synchronic linguistics. And it is this linkage that allows each to play a central role in development a coherent understanding of the other. But in my self-centered view, it is historical linguistics making the more valuable contribution here. When we explain ways in which the 'diachronic filter' can be invoked to limit the amount a machinery posited for UG, we enhance the plausibility of our overall approach. After all, 'it is in the genome' should always be an incredibly costly hypothesis: while it may someday be falsifiable, the more content we want to embed in the genome, the less acceptable the 'UG-approach' will be to biologists, and even other linguists. Historical linguistics, by uncovering the pathways of change rooted in acquisition, makes many UG-based explanations redundant, streamlining that genome-embedded system. All of linguistics benefits when we do this.

References

Anderson, S. R. (1992). *A-morphous Morphology.* Cambridge University Press.

Behaghel, O. (1932). *Deutsche Syntax: Eine geschichtliche Darstellung* (Vol. 4). Carl Winter.

Delbrück, B. (1900). *Vergleichende Syntax der indogermanischen Sprachen* (Vol. 3). Karl J. Trübner.

Hale, Mark. 1998. Diachronic Syntax. *Syntax* 1.

— — —. (1987). *Studies in the Comparative Syntax of the Oldest Indo-Iranian Languages* (Unpublished doctoral dissertation). Harvard University.

Halpern, A. L. 1992. *Topics in the Placement and Morphology of Clitics* (Unpublished doctoral dissertation). Stanford University.

Harris, Alice and Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective.* Cambridge University Press.

Keydana, Götz. 2011. Wackernagel in the Language of the Rigveda. A Reassessment. *Historische Sprachforschung* 124.1: 106-133.

Lightfoot, David. 1991. *How to Set Parameters: Arguments from language change.* MIT Press.

Roberts, Ian. 2021. *Diachronic Syntax* (*Second Edition*). Oxford University Press.

Taylor, Ann, & Susan Pintzuk. 2015. Verb Order, Object Position, and Information Status in Old English, in T. Biberauer and G. Walkden (eds.), *Syntax over Time: Lexical, Morphological, and Information-Structural Interactions.* Oxford: Oxford University Press, 318-335.

Wackernagel, J. (1920). *Vorlesungen über Syntax: mit besonderer Berücksichtigung von Griechisch, Lateinisch und Deutsch* (Vol. 1). Emil Birkhäuser.

Watkins, Calvert W. 1964. Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb. *Celtica,* 6, 1–49.

Zwicky, Arnold M. 1977. *On Clitics.* Indiana University Linguistics Club.