TOWARDS A SYNTAX OF PROTO-BASQUE

Juan Uriagereka
(UMD & UPV/EHU)

In this note I reflect on a reconstructive path for the syntax of Proto-Basque, building on the results of Gomez and Sainz (1995), who themselves work on a tradition started already by Astarloa two centuries ago, and of which Trask (1977) is an excellent example. The matter is interesting not just in itself, but also in that it may help us determine what procedures one can reliably use to figure out an extinct linguistic entity. Section one lays out some foundations of syntactic change from three different, yet related perspectives, which I call 'lenses' on syntactic change: a theory of Universal Grammar presupposed by language learners (Principles), a theory of Core Variation that involves the learner in the task of fixing open dimensions (Parameters), and a putative theory of Sociological Drift that requires learners to actually learn something and affect one another in the process. Section two relates those three 'lenses' in a model, showing how specific phenomena are best described in the interaction of these levels of linguistic reality. Section three attempts to apply the three 'lenses' in question to the reconstruction of ancient word order in proto-Basque. A conclusions-and-questions section ensues.

1. Three lenses on syntactic change

Until recently, historical linguistics was concerned with the reconstruction of languages based on morpho-phonological characteristics of words. This is not surprising. First, morpho-phonemics is where most consensual progress has taken place in linguistics, surely a consequence of the fact that morpho-phonemic evidence is the most readily available. Second, whereas it is fairly straightforward to 'integrate' source morpho-phonemic structures from their 'derived' forms after centuries of drift, it is extremely difficult to do the same for structures whose very essence is recursive; the reason is Saussurean arbitrariness at the morpho-phonemic sign. Arbitrariness entails a fixed, static structure, whose only dynamicity arises as a result of time factors. The reconstructive part has then a clear target, as it would if it were trying to find the integral of a derivative to a simple first order function. In contrast, a *combination* of signs (the *raison d'etre* of syntax/semantics) involves a moving target, especially so inasmuch as we assume no upper boundary for combinatory length. To go on with the mathematical analogy, reconstructing earlier structures in the syntactic instance would be like dealing with derivatives over time *of higher order functions*, for which integration becomes an art.

If the space of syntactic/semantic change weren't to some extent pre-established, it would be hopeless to reconstruct what went on, a bit like attempting to solve a mystery with too many uncertainties: *perhaps* there was a corpse, *perhaps* there was a murderer, and so on, ad infinitum. Luckily though, the relevant space is extremely constrained. This is most evident in the case of semantic variation, if there is any. Setting aside Worfian speculations, languages don't vary in their fundamental intentional/conceptual characterizations of linguistic reality for semantic purposes. No known language lacks (generalized) quantification, antecedence nuances, individual naming, or complex argument structuring. Languages differ on whether they divide their various conceptual spaces in thin or coarse slices, but they do not really differ on the ability to conceptualize in pretty much the same ways. Likewise, language use may differ on whether it deploys simpler or more cumbersome intentional procedures for denotations of various sorts, but reference is uniform across the species. It would thus appear that, in its essentials, human language semantics hasn't changed since the species got access to it. If so, change-wise there isn't much to account for here, and at best the issue is one of whether lexical fields shrunk or expanded. Things are tougher in the case of syntax. We know that radical word order changes took place over the millennia, or in the ability to drop arguments, or in Case/agreement, elliptical, or binding systems, etc. The question is no different from one about geographic diversity, and everyone expects that the historical stages of languages correspond to structural forms found in some of the existing languages in the world. In short, there are —there must be— dimensions of syntactic variation, the so-called parameters that input data must specify, simply because of what we witness. But again, here too we have powerful tools to constrain our linguistic search: those very cues that a human being has for unconsciously figuring out the language he or she is acquiring.

Although this ought to be a truism, it might need reiteration: Languages don't have an extra-corporeal reality, they are the products of human minds. There are thus a priori, given our understanding of human history, only three events that could be relevant to linguistic change: (i) Human evolution; (ii) human development; (iii) human interactions. The first pertains to the whole species, and as such ought to enter into a characterization of what language *is*, phylogenetically, not how it changes (whatever was prior to language is by definition proto-language, thus irrelevant to present concerns). The second event pertains to the growth of human fetuses until we can think of them as mature, in the case of language by whatever metric we take to characterize adult speech. The third event pertains to the very fact of human life, with all its nuances in the case of linguistic performance. A fortiori language change has to be defined over language acquisition and performance by humans, both ontogenetic events. The boundaries of either of those tasks must fall within biological, psychological, sociological or similar constraints that human existence may be subject to —no more, no less.

With regards to the specific topic of syntactic change, general biology, psychology or sociology have had relatively little to offer. The same isn't true about linguistics. In particular generative grammar has matured into a synthesis that could be broadly characterized as having agreed upon a few core general universal laws and a set of open dimensions of variation manifested in the various languages of the world. One version of the program is the Principles and Parameters (P&P) model; there are others within generative grammar, but all of them agree on modeling language as a universal phenomenon of the mildly-context-sensitive sort, and all of them admit that not all of language is universal. Let's think of this as a general linguistic result, calling it the P&P *program*, setting aside, as in a normal scientific enterprise that is still evolving, what is the exact content of the scientific predicates in each part of the program, the universal and the particular one. There is enough of a consensus on the major axes of that program to be able to use it as a guideline for what ought to be the boundary conditions of linguistic change. I will refer to this as *the linguistic lens* —a flexible one at this stage of our inquiry.

In addition to the linguistic lens, generative grammar has also provided us with what we may think of as a *learnability lens* and a *performance lens.* The first of these is based on the 'poverty of the stimulus' argument, the idea that it is impossible for a learner to inductively acquire a first language. David Lightfoot in his various works over the years has been the instigator of this particular lens, raising the conjecture that, if there is a point at which language may 'easily' change syntactically, that would be when it is acquired by a fresh mind. Granted, learnability considerations have not been fully worked out, and thus for instance much current works goes into elucidating the exact form of parametric triggers or cues, whether these cues are all equally accessible to a learner or some maturation is involved, whether they should be confined to root clauses, and so on. In this respect, this learnability lens ought to be seen as even more flexible than the linguistic lens, but not for that as less real: whatever turns out to be the exact path of acquisition (virtually instantaneous or maturational, for instance), what is clear is that something in this league ought to constrain language change.

The final lens I will be assumed is based on the idea that, for still unclear reasons, human groups may introduce mechanisms in their behavior to distinguish themselves from different, indeed often previous, groups. The UPenn group, with Labov and Kroch as leaders of different aspects of relevant research, have been the champions of this ultimately sociological lens. Of course, this third lens is even less solid, if only because it presupposes a detailed understanding of human societies which, for the most part, science lacks —no doubt a reflex of their complexity. In this instance I won't even venture to speculate what are areas of discussion, since I don't even know that a dominant paradigm has been established. Nonetheless, the effect of societal pressure in language change, as I will show immediately, can be logically inferred. Thus regardless of how clouded our vision may be, as of yet, through this lens, we will have to make-do with what we have and try to go from there towards a reconstruction that, otherwise, would be impossible.

Corresponding to these three lenses —a point about linguistic methodology— I will suggest in the next section that —somewhat surprisingly as that will be a point about linguistic reality— human language involves three separate layers, each progressively less robust in the sense that it is less a consequence of established natural principles. I will lay out these ideas in more familiar terms there, but in order to be clear about the need for the sociological lens, I want to note that one neutral way to understand the layered structures of language is in terms of the expression $uf(p_i)$, where u stands for 'universal grammar', a constant biological specification, and $f(p_i)$ is a function characterizing a core grammar as standing in some sort of relation with regards to individual performances, p_i, where the sub-index i ranges over human individuals. (Note, so far that only gives us two layers: u and f). Simply because human lives differ, two individuals sharing the very same linguistic competence will never exhibit the exact same performance. In most instances we presume these individual differences not to coalesce into any significant trend. However, peripheral variants can surface if, for poorly understood reasons, particular surface phenomena do coalesce somehow. Put differently, it is trivially the case that, for any given value of i (for any person), the derivative of $f(p_i)$ over time (their actual linguistic performance) is (uninterestingly) different. At the same time, and intriguingly so, if we look not so much as each value of i in the function above, but rather at a more abstract cut on p_i intuitively dealing with some sub-population s (for $s = \Sigma^n_{i=1} p_i$), one cannot fail to notice, as it were at some macroscopic level, that patterns may emerge on structural possibilities that are not fully specified by a core grammar. Then $f(p)$ becomes a more abstract function over, in effect, population dynamics (i. e. $f(\Sigma^n_{I=1} s_I)$, for s the sum of individual performances just alluded to). I have little to say about these dynamics as such, but I want to emphasize that, aside from arguably important on their own right for one sort of (smooth) linguistic change, they are also presupposed in the other approach to (radical) language change that we have mentioned.

For suppose we idealize language as being a constant k, once acquired, within an individual; the derivative of k is zero. If this were all there is to it, language wouldn't change at all. But of course human life is short, and new generations also acquire k. Now say the acquisition process is some function $uf(\delta)$, where u stands for 'universal grammar' and δ for 'input data'. We know by considering the entire human population that the output of f, though still arguably constant for a given individual, is not constant for the species. Which is to say that the derivative of f over time may well be interesting. This can't be because of u, a constant; however, δ is a reasonable culprit: input data is a priori subject to fluctuation sources. One limiting condition of this situation emerges if generation x does not provide data δ to generation $x+1$, but some outside force does. This may result in language change, but of course of a purely external, thus sociological, nature. I don't mean to imply that this is trivial from the point of view of the performance lens, quite the opposite: Kroch has argued (e.g. in 2001) that some of Lightfoot's claims about, in particular, the evolution of English are best understood in these sociological terms. However, this situation is not very interesting from the point of view of the learnability lens, since there need not be any curious property of the learner that would account for this totally expected result: different input δ yields different output $uf(\delta)$. A more remarkable situation would emerge, though, if from essentially the same input δ_I we were to obtain different output $uf(\delta_I)$. Evidently, this is impossible if not qualified; I have added the individuation subscript I to δ_I precisely to indicate that, although across generations δ may be essentially the same *kind* of data, individual variations glued together into population drifts may we be very significant in making δ_I different, from the purposes of acquiring a grammar, from

δ_J (meaning by *I* and *J* here simply input data from different generations of speakers of the same 'language').

In other words, inasmuch as grammars subjected to accepted population dynamics (notated as '>>>') of the form in (1)

(1) $\Sigma^n_{I=1} (\Sigma^n_{i=1} p_i)_I >>> \Sigma^n_{J=1} (\Sigma^n_{j=1} p_j)_J$

yield peripheral data drifts $\delta_I \rightarrow \delta_J$, then it must be that these drifts, as Lightfoot argues, end up being responsible for changes in core grammar. That in turn suggests that the function $uf(p_i)$ that we started this discussion with must be understood as $uf_{vt}(p_i)$, where f is a complex function that at least in part relates parametric values v to what we may think of as peripheral trainings t. It is not my intention in this note to clarify just how it is that, in effect, linguistic growth (parameter values) and linguistic learning (peripheral trainings) relate to one another. The point is, simply, that it must be in this way that we resolve the, otherwise absurd, situation for linguistic change that Lightfoot hypothesized, leading to parameter value resetting of the catastrophic sort.

2. The Assumed Model

Those abstract considerations can be made concrete within the P&P model. Let's first examine, in that respect, instances of the Polysynthesis parameter (in the sense of Mark Baker's), which gives different values for English and Basque: while the latter case-marks the main verbal arguments (subject ergative, object absolutive, indirect object dative) and exhibits them as agreement elements in the verbal auxiliary, the former does neither. Correspondingly, English disallows pro-drop (2a), presents a fixed word-order, and allows extractions from nominal objects (4a), whereas Basque can drop any of the main verbal arguments (2b), allows any of the orders in (3), and disallows extractions from nominal objects (4b); 'pied-piping' extractions are fine in both languages:

(2) a. *(I) love *(Lucy).
    b. maite nuen 'I love him/her/it'
       Love I-AUX-III

(3) a. Nik maite nuen Lucy. 'I love Lucy.' (OK in English)
    b. Lucy nik maite nuen. 'Lucy I love.' (OK in English only with topicalization)
    c. Nik Lucy maite nuen. (Permutations OK too) 'I Lucy love.' (etc. all * in English)

(4) a. Who has John seen [pictures of t]?
    b. *Noren ikusi ditu [t argazkiak] Jonek ('Of whom has John seen pictures?')
       who-gen see III-AUX-III pictures-abs Jon-erg

(5) a. [Pictures of whom] has John seen t?
    b. [Noren argazkiak] t ikusi ditu Jonek 'Pictures of whom has John seen?'
       who-gen pictures-abs see III-AUX-III Jon-erg

This is the expected situation in parametric choices, which typically correspond to low-level morphological facts (case, agreement) and have vast consequences for the grammar at large (in terms of syntactic conditions of all sorts).

Importantly, neither the Basque nor the English value of that particular parameter seem to be more marked than the other. Suppose we think of markedness in classical Pāninian terms, via the Subset Principle. Whenever we find situations whereby a set of structures in language L is a subset of a larger class of structures in language L', we assume that the language acquisition device (LAD) must hypothesize that it is acquiring language L unless presented with direct positive evidence for a structure in the superset. Had the LAD hypothesized, in the absence of such confirmation, that it was learning the language corresponding to the larger set, the only way it could retreat from a mistaken assumption is by way of analyzing negative data. Which is the larger set of structures related to (and therefore, which can set) this particular parameter —the Basque or the English one? If we go with the evidence in (3), the Basque structure is the superset (as more grammatical combinations with the same words and identical grammatical relations are possible in this language); but if we go with the evidence in (4) and (5), the opposite is the case. So the LAD cannot decide which one is the unmarked option. That is not problematic, so long as robust positive evidence exists for each option of the parameter. The evidence is overwhelming if we allow the child to unconsciously analyze the syntactic properties of sentences. An English speaking child cannot help but find the strongly configurational language that structures of the sort in (1a) imply. There are very small departures from this kind of evidence, such as the exotic topicalized (3b), with special intonation and (for most people) restricted to main clauses. But the evidence for the Basque child is equally strong: every Basque sentence presents an inflected auxiliary and either dropped arguments or overtly case-marked ones, both hall-marks of the opposite setting in the Polysynthesis parameter. In the presence of robust evidence for both settings, learning either is trivial, and there is no logical need to postulate an unmarked option which could not be learned on the basis of only positive data.

Not all parametric situations are like that. Compare languages, like Spanish, which allow clitic-doubling, and languages like English that don't, as in (6):

(6) a. Juan la está viendo (a María) (Juan is seeing María)
Juan 'er is seeing to María
b. John is seeing'er (*Mary)

Let's call whatever is involved in this difference the Clitic Parameter. Inasmuch as doubling is optional (6a), the set of structures it allows is a super-set of the set of structures associated to absence of doubling (6b). One could argue that, just as (4b) is impossible in languages with object agreement, so too is (7) impossible:

(7) *De quién las está viendo [a amigas t] ('Who is s/he seeing friends?')
of whom 'em is seeing to friends

However, that would be an unfair comparison. This is because (8a) is as grammatical in Spanish, crucially without a clitic, as its English version in (8b) is:

(8) a. ? De quién está viendo [amigas t] ('Who is s/he seeing friends of?')
of whom is seeing friends
b. ? Who is s/he seeing [friends of]?

Note that one cannot build, in a language with obligatory agreement like Basque, a version of (8a) *without* agreement. That is one of the central differences between clitics and agreement markers: the latter are not dropped. Note the consequence of this state of affairs: a grammatical version of (7) exists in Spanish, so in this instance there is arguably no structure that the English version of the Clitic Parameter allows which Spanish doesn't, and hence English is, in this domain, a genuine subset of Spanish.

The subset situation just described doesn't entail that Spanish (the marked option) will take longer for a child to learn. If as we did in the previous paragraph, we allow the child access to second-order grammatical descriptions of available data (as first suggested by Dresher 1998), then it may well be that the evidence cueing even the marked option of the parameter happens to be readily available for a child to use. That highlights the difference between the current P&P model and a more traditional proposal (such as the *Aspects* model of Chomsky 1965) in which learners always compare grammars in terms of the first-order linguistic structures (sentence tokens) that they license. In genuine parametric terms, a set comparison of the sort alluded to may be relevant only in situations of a learning conflict, where higher-order evidence leads to ambiguous analyses. Still, the scenario outlined in the previous paragraph is important in principle, and arguably central in fact for situations of successful language change. After all, there may well be a historical stage in which a learner lacks a strong way of directly deciding whether the language being acquired involves genuine syntactic clitics (as opposed to merely reduced phonological forms, which all languages have in casual speech). In scenarios where that grammatical ambiguity ensues, it is only natural for children to resort to first-order data analyses, going with the subset condition in the absence of direct observation of grammatical tokens of the superset sort.

But there has to be more to linguistic differences than mere parametric settings, whether they are of the subset sort or not. Chomsky is very explicit in 1981 about the role of idealization and how that relates to the notions we are considering. He says that:

> . . . what are called 'languages' or 'dialects' or even 'idiolects' will [not conform] to the systems determined by fixing the parameters of UG. . . each actual 'language' will incorporate a periphery of borrowings, historical residues, inventions, and so on, which we can hardly expect to —and indeed would not want to— incorporate within a principled theory of UG.

Nonetheless, Chomsky also emphasizes that 'outside the domain of core grammar we do not expect to find chaos. Marked structures have to be learned on the basis of slender evidence too, so there should be further structure to the system'. While that is reasonable, it is difficult to pin down the nature of that further structure. Chomsky says:

> . . . we assume that the child approaches the task [of language acquisition] equipped with UG and an associated theory of markedness that serves two functions: it imposes a preference structure on the parameters of UG, and it permits the extension of core grammar to a marked periphery.

The first of these functions is obvious, and has been discussed above; the second one is less so. It is trivial if what we mean by 'periphery' is simply a core grammar with marked options set. However, there is no reason why all 'borrowings, historical residues, inventions, and so on' (where the 'so on' could be pretty vast) ought to be that clean.

Consider a situation that arises with questions involving the operator *why*. First, it should be said that in many languages regular argumental operators such as *who* trigger, upon fronting an information question, an ancillary verb movement, involving an auxiliary (as in the English (8a)) or the main verb itself (as in the Spanish (8b)):

```
(8)  a.  [Who [has [John [ t [seen t ]]]]]? (cf. *Who John has seen t?)
           ^      ^_ _ _ _ _/        /
           |_ _ _ _ _ _ _ _ _ _ _ _/
```

```
     b.  ¿[A quién [vio [Pedro [t [  t   t]]]]] (cf. *¿A quién Pedro vio?)
           to whom saw Pedro /^   /   /
               ^      ^_ _ _ _ _/ |_/   /
               |_ _ _ _ _ _ _ _ _ _ _/
```

Why the verb movement is required has been a topic of much discussion. Suffice it to say that a sentence boundary (technically, a CP) is (in most languages) a barrier, and the 'barrierhood' of this category disappears if the CP is in construction with a lexical head:

(9) *L(exical)-marking Convention*
XP is not a barrier if it is in construction with a lexical head Y.

Where X is in construction with Y if X=Y or X is selected by Y.

(The disjunction in (9) disappears if categorial identity is a trivial sub-case of selection). By the L-marking Convention (LC), an embedded CP does not require verbal inversion of the sort above. Thus, compare (8a) to (10a), which does not involve verb movement:

(10) a. [I wonder [$_{CP}$ who [John [has [seen t]]]]]
b. ... [$_{CP}$ C [$_{IP}$ John [has [$_{VP}$ seen who]]]]

(10b) goes back in the derivation to the point prior to the displacement of *who* (simplifying the structure for expository purposes). Suppose all XPs along the way are potential barriers for this displacement, keeping in mind that an XP is just the maximal projection of X at some derivational step, whatever X may be. VP is in construction with a lexical head, namely its own; as a result VP is not a barrier for the movement of *who*. Of course, by this sort of reasoning, all categories headed by a substantive head will not be barriers to displacement. If the T head of IP also counts as substantive, then IP will not be a barrier either. Finally, how about CP? Its head is the abstract C, thus clearly not substantive. But is it in construction with a selecting element? It is if the CP is selected by *wonder*. Then CP is not a barrier either, though not for intrinsic reasons (its own head), but rather by way of its contextual properties (the selector that licenses the structure in the derivation).

This is the step missing in (8a), where nothing selects the relevant CP. As a result, the only way CP will cease to be a barrier is if it incorporates a neighboring lexical head, in which case the combined projection will be in construction with an appropriate L-marker. That is precisely what head movement achieves:

(11) [$_{CP}$ has-C [$_{IP}$ John [t [$_{VP}$ seen who]]]]
^_ _ _ _ _ _ _/

Observe that the displacement of the verb does not cross the CP, but adjoins instead to the head C, and is thus legitimate. In contrast, the displaced *who* must cross CP, but this category is no longer a barrier after verb movement, in the manner indicated. The reasoning is rounded up by the assumption that the mechanism in (11) is, in some relevant sense, costly, which is why the grammar does not undertake it if not necessary; hence inversion in the circumstances in (10) is not acceptable.

Assuming a system along those lines, the question arises about structures involving adjunction to CP itself, which thus should not cross this element to begin with. As Rizzi (1990) has indicated, this situation arises for causal modifiers, and therefore for corresponding questions involving the operator *why*. The logic is clear. *Why* questions involving one level of embedding should not trigger verb preposing as discussed above; however, *why* questions involving two levels of embedding should. In other words, (12a) should be grammatical, alongside with (12d), while both (12b) and (12c) should be bad:

(12) a. Why [$_{CP}$ John has seen Mary] t? ☹
b. * Why has [$_{CP}$ John seen Mary] t? ☹
c. * Why [$_{CP}$ you have thought [$_{CP}$ John has seen Mary] t] ?
d. Why have [$_{CP}$ you thought [$_{CP}$ John has seen Mary] t] ?

In (12a) *why* does not have to cross CP, thus it should be unnecessary to move *has* as in (12b) to void the barrierhood of this CP. In contrast, although *why* in (12c) does not have to cross the embedded CP it modifies, it does have to move across the matrix CP in its displacement to the clausal periphery; hence this time ancillary verbal displacement to the C head is justified. Speaker judgements about (12c) and (12d) accord with the theoretical prediction. However, judgments about (12a) and (12b) are backwards (☹ in (12)).

That is not so with English-speaking children who, as Crain and Thornton (1998) show, provide just the judgments in (12). Spanish too presents almost the same paradigm:

(13) a. ¿Por qué [$_{CP}$ Juan vio a María] t? 'Why Juan saw María?'
why John saw to María]
b. ¿Por qué vio [$_{CP}$ Juan a María] t? 'Why did Juan see María?'
why saw John to María]
c. *¿Por qué [$_{CP}$ tú pensaste que [$_{CP}$ Juan vio a María] t]? 'Why you thought that
why you thought that John saw to María] John saw María?'
d. ¿Por qué pensaste [$_{CP}$ tú que [$_{CP}$ Juan vio a María] t]? 'Why did you think that
why thought you that John saw to María John saw María?'

The parallel is not total, since both (13a) (involving no verbal displacement) and (13b) (involving it) are possible. Nonetheless, facts are similar enough for Crain and Thornton to make their point: in their view of things, children acquiring a language L cannot make a 'mistake' which does not correspond to a parametric option in some other language L' (their Continuity Hypothesis). Suppose they are correct, a question remains: what is responsible for the English pattern in (12) (or related to this question, why is the Spanish (13b) —involving what looks like a costly and unnecessary option— also an option alongside the predicted (13a))? Actually, vernacular versions of English present the pattern in (12) as well, and upon closer examination, the Spanish (13a) belongs to a more relaxed register than (13b). It is possible that the verb preposing in (12b) or (13b) is one of those 'peripheral inventions' that Chomsky was referring to, achieved on analogy with instances of verb preposing where it is needed in order to eliminate a barrier by way of the LC in (9). That would explain why children continue to use the pattern predicted by the theory sometimes into early puberty (always beyond normal stages of acquisition), as do 'uneducated' speakers. This is not to say, of course, that the prestige or educated adult pattern is not psychologically real. It is very possible, however, that its acquisition constitutes a genuine instance of learning, and as such is entirely different from whatever is involved in more elementary parameters which I have likened to growing.

As with other instances of linguistic diversity, we do not know for sure that this type of analysis of *why* questions is ultimately correct. As Chomsky asked in (1981):

> How do we delimit the domain of . . . marked periphery? . . . [E]vidence from language acquisition would be useful [, but is] insufficient to provide much insight. . . We are therefore compelled to rely heavily on grammar-internal considerations and comparative evidence, that is, on the possibilities of constructing a reasonable theory of UG and considering its explanatory power in a variety of language types.

If we base our conclusions on what is a reasonable theory of UG, we have something to say about, in particular, verb preposing for argumental questions, but not so much for similar processes involving causal modification questions. That said, it seems significant that late acquisition of verb preposing in *why* environments, or its absence in spontaneous speech in many instances, should correlate with the theoretical conclusion. This strongly suggests that there is room for a significant, systematic Periphery of a sort that seems profoundly different from whatever is involved in the constitution of I-language.

The diagram in Figure I is intended to convey this fundamental difference (growth vs. learning) between the combinatorial systems of language and those which are based on an ultimately sociological exchange. In the terms discussed in the first section, for a complex function $uf_{vt}(p_i)$ that relates parametric values v to peripheral trainings t, UG in the adjacent diagram is the u constant in the function; the core in the diagram is $uf_v\,(p_i)$ in the function, which eventually becomes an I-language when *all* the v (parametric) values are set through input data δ for a given individual i. Note that the transition from the Core to the I-language in Figure I is

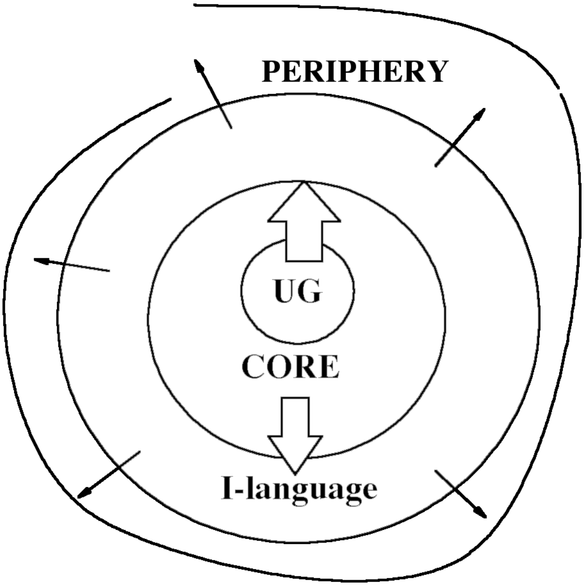

Figure I

From I-language to the Periphery

not signaled in the function as such; this is because it is not an architectural issue, but an acquisition one: data for the Core, if available, are direct second-order analyses of δ, while data to entirely set an I-language may need to involve indirect first-order analyses of δ in subset situations. (In principle, Core and I-language may end up coinciding if either there is no second order δ for the child to set or no sub-set situations where first order δ become relevant). Finally, the periphery in the diagram is $uf_{vt}(p_i)$ (the whole function), which becomes what we normally understand as a language when all the *t* (learning) trainings are determined, also through input data δ for *i.*

A model of grammar of the sort in Figure I may undergo interesting patterns of change. Suppose a sociological drift takes place in some peripheral construction (e.g. an augmentation of displacement processes of the sort in (13) for expressive purposes). That, in itself —if it becomes stable enough to last— will of course constitute a *bona fide* instance of linguistic change (and it would presumably present the characteristic 'S'-shaped curves of such transitions). For my purposes here, however, the most interesting consequence that such an S-shaped curve may end up having is with regards to the data, δ. A learner doesn't consciously discriminate between what I have called second-order or first-order δ, or for that matter δ for the purposes of $uf_v(p_i)$ and δ for the purposes of $uf_{vt}(p_i)$ as a whole. Data are data. However, we already saw that the first and second-order interpretation of δ is quite different: in one instance sets of tokens are relevant, whereas in the other, instead, what counts are grammatical structures. Similarly, what we may think of as δ_c (for core settings) are interpreted by the system differently from what can be regarded as

δ_p (for peripheral learning). In principle, all of this is no different from the use a given organism may have for a key protein first in a purely genetic mechanism, later on in the life of the organism in epigenetic processes, and eventually in metabolic exchange. In our terms, the very same δ is first used in setting core parametric values (first of the cueing sort, next in terms of set evaluation), and eventually in whatever mechanics are involved in establishing peripheral patterns. This hierarchy is crucial in the dynamics for language change. An innocent-looking drift in δ (an S-type process of the sociological sort) cannot be kept from affecting those aspects of the data that may tilt the $uf_v(p_i)$ part of the equation. For instance, a simple frequency change of the peripheral sort can, no matter how rarely, affect the first-order subset conditions for a learner to set the paradigmatic value(s) of a subset dimension like the Clitic Parameter. A change of the latter sort may in turn, again if only sporadically, imply the emergence of further sorts of evidence which, when analyzed at a second-order, result in different conditions for a learner to set a core dimension like the Polysynthesis Parameter. That's Lightfoot's rationale. If this sort of, in his own words, *catastrophic* sort of change may occur with any probability higher than zero, it will result in a non-linear curve, signaling a drastic syntactic change within a few generations (one or two, in the limit).

3. Reconstructing Basque Syntax from Auxiliary Morphology

In understanding what I have to say about the possible course of change from Proto-Basque to Basque, it is important to realize that a hierarchy is being assumed:

(14) Peripheral change > Subset change > Core change

This is the most realistic course of radical linguistic change: a disturbance emerges in the periphery of a language which manages to cascade through interesting consequences for the first-order analysis of δ, the data available to a language learner, and eventually δ's second-order analysis. This course of action is not necessary: a disturbance in the Periphery may simply stay there, and even if it manages to affect an I-language, it may never trickle down to Core parametric options. Nonetheless, *if* a radical language change is to take place, logic dictates that it proceed in the manner assumed in (14) —unless it is directly imposed by outside forces (e.g. a foreign invasion or forced enslaving dramatically changing input data for a generation of children). It could be argued that the hierarchy in (14) could 'skip' a step. Logically that is possible, although it is hard to imagine what sort of trivial change in the periphery —a first-order one— may be such that it, and it alone (without a subsequent change in the I-language), can affect the way in which data are analyzed at a second-order at the core. If this situation emerges it requires careful argumentation. At the same time, and putting aside irrelevant language change of an external sort, the present logic does not expect a *direct* change to occur either in the I-language or in the Core. While it is possible for a generation *of adults*, following societal pressures, to change their first-order data, it is impossible for them to engage in more elaborate manipulations, as it is impossible for children to change anything at all: they simply analyze data. In other words, while the cause for radical language change may be children, the trigger must be adults (because they change sociological patterns, or they invade/enslave one another). Differently put, the linguistic lens forces us to look at data both from the performance and the acquisition lenses, in that order.

That said, the main idea from the Gomez and Sainz paper (henceforth G&S) that I want to address here is their claim that 'Basque had a V1 order at a certain stage… This situation was responsible for the reanalysis of connectives as person markers and ultimately for the phenomenon of ergative displacement' (section 6.3). The methodology that G&S uses to reach this conclusion is very sound. First they carefully study the constituent elements of finite verb forms in Basque (root, person and number affixes, and event markers for mood, tense and aspect), together with the conditions of split ergativity witnessed in the language, from the earliest (sixteenth century) surviving texts. Next, they take seriously the traditional conjecture that, in particular, personal agreement markers have their historical origin as incorporated personal pronouns. Then they proceed to analyze the more complex (and in their plausible view, later) origin of number forms, finally discussing event-related markers. Finally, they ask the question that interests me here: how is it possible for a finite verb form to incorporate all those elements?

A by now familiar dictum by Givon in (1973) asserts that 'today's morphology is yesterday's syntax'. The intuition behind this bold assertion is that morphology acts as a reflex of syntax by somehow 'grammaticalizing' syntactic patterns. Of course, if one doesn't believe in the extra corporeal reality of grammars, 'grammaticalizations' ought to be made compatible with the sort of system sketched in the previous section and summarized in (14) above. In broad terms, the obvious intuition to pursue, much in the spirit of Clark and Roberts (1993), is that 'grammaticalization' is performed by language learners, upon simply reanalyzing originally periphrastic sequences as word-level units. For that a couple of things must happen: (i) that the periphrastic sequences are stable enough for a language learner to analyze them as (presumably simpler) units; and (ii) that they were not stable in the particular sense in (i) prior to the point of misanalysis. In addition, for us linguists to reconstruct what went on in history, a third desideratum must be met: (iii) that morphological rules haven't destroyed the evidence of prior syntax. Let's consider each of these important points in turn in order to analyze the plausibility of claiming that Basque was a V1 language at the point of creating its finite verb system.

Intuitively, desideratum (i) asserts something as simple as claiming that the noun, say, *pickpocket*, emerges in analyzing as a unit the frequently co-occurring verb *pick* and noun *pocket*. But needless to say even there we start seeing an issue: it obviously isn't the case that *pick* and *pocket* always co-occur; they just have co-occurred frequently enough in the (relatively recent) history of English for someone to put them together, with the innovation somehow sticking. Granted, this trivial example involves substantive lexical items, whereas what interests us here is the formation of verbal *paradigms*. But still, the *pickpocket* example highlights a point of principle: just how frequently must grammatical form *x* co-occur with grammatical form *y* for them to be turned into something that learners acquire as the unit *x-y*? I don't know of any answer to that question, but it seems to me rather easy to prove that the answer cannot be 'all the time'. Take for instance the origin of complementizer *whether* in English, which is known to have emerged from the morphological merger of *which* and *either*. Clearly it could not have been that these two words were *always* coupled together, or they would never have remained separate in the language. The same thing can be said about complementizer *como* in Spanish, originating from the vernacular Latin *quo modo*; plainly, Spanish has retained separate words originating from *quod* and from *modo*. The examples that come to mind are too numerous to mention, so the conclusion seems inescapable:

(15) Grammaticalization *g* does not require constant co-occurrence of its constituents.

That moves me to desideratum (ii). One other reason to believe the claim in (15) is that a grammaticalization *g* takes place at some point *t* in time. Simply put, if the conditions for *g* had been met prior to *t*, then *g* should have occurred prior to *t* as well. But we don't expect things to be like that in any dynamic system: transitions take place only when certain limiting conditions are reached —the proverbial straw that breaks the camel's back. If so, what we expect in grammaticalizations is some sloping curve that can create the conditions for a change, which *then* takes place. This makes sense from the point of view of language learners receiving data δ, assuming that:

(16) (i) A language learner acts economically.
(ii) It is more economical to pre-compile a symbol sequence $<x, y>$ as a word w than as separate symbols x, y.

From the point of view of the grammaticalization, then, we should expect that:

(17) Grammaticalization *g* is a transition at a time *t* from a syntactic state *s* to a morphological state *m* when the δ evidence for learners allows for their most economical analysis whereby $s > m$ (i.e. morphology trumps syntax).

Prior to *t*, δ evidence is just too scarce for the desired analysis from the learner. The transition from *s* to *m* may be more or less drastic (linear or non-linear) depending on where in the hierarchy in (14) the change in question falls: whether it is Peripheral (as possibly lexical changes of the *pickpocket* sort are) or Paradigmatic (as ought to be the case with both Subset —e.g. expansion/reduction of paradigms —and Core changes —for instance in value for the Polysynthesis Parameter). Regardless, given this state of affairs, it cannot be the case that a linguistic change happens at either limiting situation (total absence or total presence) for δ, unless there is some sort of drastic invasion or some such thing that directly changes the nature of δ for a generation of learners.

Finally, desideratum (iii) above is methodological. Any unqualified version of Givon's grammaticalization claim is bound to be false, for the simple reason that we know (as amply shown for instance in Anderson 1992) that morphology too has rules. That said, all we can do is hope for the best. If in particular Baker is right in (1988) and subsequent works, morphology, or more narrowly inflectional morphology, is of the sort that mirrors the syntax that goes into it, in fact even synchronically. That is the strongest sort of condition one can hope for, if G&S is to be taken seriously, for that work is essentially claiming that the synchronic inflectional morphology of the Basque finite verb wears the syntax it came from on its sleeve. If so, that is a wonderful fossil record to reconstruct the origins of Basque syntax from, mainly because we are so lucky that the Basque verb happens to be so rich in its morphology. At the same time, even if one has perfectly fossilized remains of anything, one still needs good reconstructive science to figure out what went on. This entire note can be seen in that light, as an attempt to use the linguistic, learnability and performance lenses to help us test the reconstruction of Basque syntax, and in particular the claim that it was verb initial.

The reason G&S wants the Basque verb to be verb initial —involving a verb movement very much along the lines of (8) above— is that, in this context, around the verb root 'a number of clitic-like elements [of the sort in (6) above] were attached in such a way that a kind of verbal complex was created' (cf. the examples in (2) to (5) above, where the auxiliary shows rich personal morphology in a positive value for the Polysynthesis Parameter). Why couldn't these elements be attached if the verb were any lower? The tacit assumption is that in lower contexts there wouldn't be enough structural height for the verb to be systematically close to the would-be morphemes. The point can be illustrated in the following diagram (word order is orthogonal to the point being made). Suppose a syntactic derivation for a transitive structure involves (18):

```
(18)    CP
       / \
      C   TP
          / \
         DP  T'
             / \
            T   vP
                / \
               DP  v'
                   / \
                  DP  v'
                      / \
                     v   VP
                         / \
                        V   DP
```

(18) is the standard Minimalist phrase-marker, but any other one with similar properties would do for our purposes. The key is that the verb starts its derivational life 'low', associated to the direct object, its thematic subject being associated to an auxiliary verbal element *v*. Verbal arguments may displace to higher *v*P or TP specifiers for Case/agreement purposes, while the verb itself may move, head-to-head, all the way up to the C projection. Whenever I say operations *may* take place, it is because a parameter could be involved (thus the English verb apparently doesn't move as high within the clausal skeleton as the Spanish verb, etc.). The key point is this. The traditional intuition has always been that verbal (personal) agreement emerges in history from cliticization of D argumental elements. If Baker is generally right, this presupposes the process of Incorporation, which must satisfy the following principle:

(19) *Principle of Incorporation*
An incorporation process from *x* to *y* takes place only if *y* c-commands *x*.

If (19) holds, then it is immediately true that incorporation of anything to a verb presupposes the verb being 'high enough' (in c-command terms) for the incorporation to take place; for subject and object clitics, that must be C. G&S generalizes this result to other verbal morphemes, but the same point holds: *whatever* incorporates to a verb presupposes a structurally high verb. Hence the verb must be in C to succeed in the task.

I agree with this result. The only ways to prove it wrong would be either to show that morphology in Basque doesn't obey mirror conditions of the sort Baker analyzed, or else to deny (19) altogether, in which case no strong conclusion about verbal height directly ensues (a possibility explored for instance in Di Sciullo and Williams 1988). However, I find the explanation interesting enough, and the theory behind it sufficiently strong, not to challenge that premise. What I want to focus on, however, is the issue of whether in order to have the verb sufficiently high to be able to host incorporated elements, it must be in C *all the time* —in other words, whether as they claim the grammatical order of Basque at the time must have been VSO (or putatively VOS).

That claim is radical, especially in light of the fact that modern Basque is almost certainly underlyingly SOV or perhaps SVO (in part depending on whether the conclusions in Kayne (1994) are tenable and whether, in contrast, there may exist a Head Parameter determining the order of a verb and its direct object). To realize this, compare the assertion that Proto-Basque was VSO with Trask's (1977) contention that Basque was originally SVO and it became SOV as a result of the emergence of an ergative system in the language. A priori, that position is easier to justify, from the viewpoint of the linguistic lens. In particular, this is an expected change, both typologically and synchronically: as Trask (1979) already observes, SVO languages are never ergative. But going from a VSO order to a SOV order is no trivial task. To put this in perspective, let's consider first what may be involved in the different verbal positions (higher or lower in the clausal skeleton in (18)). In a nutshell, if it were possible for it, a verb would always stay lowest in the clausal skeleton, thus yielding SOV or SVO orders (depending on whether the object itself can appear on either side). This is because such a solution is the most economical (involving less displacement), which accords well with the fact that newly formed creole languages —by hypothesis starting with default parametric settings in the absence of grammaticalized morphological markings— are typically SVO (Bickerton 1990). So why does the verb ever move higher, then, to the inflectional projections of T, C or others? All other things being equal, to implement some form of morphological checking, as G&S correctly assumes following Chomsky (1995). If so, it would be sufficient for a child to misanalyze the need for morphological checking at a higher site (C, for the VSO order) for the economy strategies of grammar to yield a lower movement site (if any) for the verb. That's actually good news for the G&S hypothesis: all that it requires is to supplement the already argued for idea with the addendum that, at some time in the historical change from Proto-Basque to Basque, the language somehow lost morphological 'strength' in the C (presumably also T) areas, which resulted in the verb not having any business going higher that it needed to. That does model the correct change —*if*, that is, the morphological weakening can be substantiated.

Here an interesting tension arises. Let's call transition time t_0 the one whereby personal clitics are integrated into the verb as agreement markers —and keep in mind that this involves a setting of the Core Polysynthesis Parameter. As per generalization (17), there must have existed a time t_{-1}, prior t_0, whereby the periphrastic construction(s) that give raise to the morphological reanalysis haven't been reanalyzed yet. And in turn there must have been a different time t_{-2} where either the periphrastic constructions themselves hadn't emerged yet because the relevant clitics weren't in place, or else the verb wasn't high enough to be in the sort of position that it, by hypothesis, occupies at t_{-1}, which is what eventually allows the reanalysis at t_0. Certainly at time t_{-1}, and arguably at time t_{-2} (if the periphrastic constructions hadn't emerged), the verb must have already been high enough to be able to result in the t_0 reanalysis, which in turn entails that Proto-Basque at the time must have had 'strong' morphology in C to trigger the costly verbal move to that site. The questions then are these: if throughout a period $t_{-2/0}$ Proto-Basque had strong morphology in C, why did it eventually lose it and what was the time t_1 when that happened? Note that as per the logic of the hierarchy in (14), the $t_{-2/0}$ period must have been either externally accelerated, or else a rather long one (to allow for the integration of a peripheral drift into a Core parametric option). That result holds all the more so if, as G&S correctly argues, not all the integrations in the Basque verbal morphology are co-temporaneous, which presupposes a *period* whereby the verb is high enough to incorporate (different) personal markers, number markers, and event-related markers. The same conclusion can be reached about the $t_{0/1}$ interval (the length of this period depending in part on whether Verb Movement parameters are either Subset or Core ones). This is then the crux of my question: the interesting G&S proposal is missing one non-trivial component to be fully accepted. In the period $t_{-1/1}$ (possibly $t_{-2/1}$) Basque went from lacking 'strong' morphology in C, to having it, to losing it yet again. That plainly requires either of two very intriguing premises: (i) there may have been (an) unexpected outside influence(s) in the verbal syntax of Proto-Basque, or (ii) the time period $t_{-2/1}$ is extraordinarily long —in the order of millennia, not centuries.

A less drastic alternative is that proto-Basque wasn't a V1 language, after all. Notice that all the G&S requires, especially given desideratum (i) above (which merely requires *enough stability* in the desired periphrastic forms), is that the tensed verb be in C with sufficient frequency for learners to be "fooled" into reanalyzing periphrastic forms as synthetic. That isn't difficult. Consider the well-known phenomenon of *galdegaia* in modern Basque, or the requirement that focused elements be left-adjecent to verbal forms. Concentrate in particular in the Northern, root-clause-driven form of the phenomenon, which is also the most archaic (and is preserved in the Southern varieties only in emphatic and negative contexts). Relevant examples have the shape in (20):

(20) Lucy ba nuen nik maite!
Lucy-abs indeed III-Aux-I I-erg love
'It is Lucy who I do love!'

Here the tensed verb is high, where G&S expects it —in a language which is underlyingly SOV. This is important, because if this is a domain that is readily accessible to language learners, and it appears with sufficient frequency for them to consider economical the appropriate grammaticalization, then we have modeled the G&S results, albeit without having made any difficult-to-justify charges regarding the syntax of Proto-Basque.

It is quite remarkable that the situation in (20) should be specific of main clauses, and in that respect counts as what Lightfoot (1991) and elsewhere calls 'Degree-0'. Simply put, whereas one could have a main clause without an embedded one, the latter is impossible; hence the frequency of main clauses is overwhelmingly higher (in the vicinity of 98%) than the frequency of embedded clauses. That statistical fact alone makes main clauses the most important repository of systematic data for children. Now crucially, main clauses typically present so-called root transformations, of the sort sketched in (8) vs. (10) above, which Laka (1990) persuasively argued is essentially of the same sort as (20). Lightfoot has shown that children typically induce catastrophic language changes in the direction of whatever surface forms they can detect from main clauses, reanalyzing I-language parameters on the basis of this sort of evidence. In (1999) Lightfoot shows that it is enough for children to obtain data patterns in the vicinity of 30% of a regularity (vis-à-vis absence of that regularity in comparable contexts) for them to be lead astray into reanalyzing particular input data as setting a parameter value that is different from the target one. Extrapolating from this appreciation in the realm that Lightfoot was stydying, it might have been enough for children acquiring Proto-Basque to receive data of the form in (20) with a 30% frequency, or any such relatively low frequency (vis-à-vis similar constructions, albeit of a non-emphatic sort, whereby the verb doesn't appear in the left periphery) for them to reset the Polysynthesis parameter in the direction that G&S expects.

It may be argued that learners of Proto-Basque wouldn't have reset that Core parameter if they could analyze data like (20) as emphatic, focused, or some such thing, thereby unconsciously realizing that the order in (20) isn't basic. However, Crain and Thornton (1998) have shown that, whereas they are extremely sensitive to core syntactic and semantic information, young children are not particularly well attuned to general pragmatic information. Emphasis and focus involve a complex interplay of syntactic, semantic and pragmatic information, so it is not surprising if a child hearing a sentence like (20) may be more sensitive to its syntax/semantics than its pragmatics, as a result of which the possibility of misnalysis is in principle there. What remains to be seen is whether one has a way of determining the frequency of sentences of the sort in (20) within a normal sample of primary linguistic data. We won't be able to search for an answer on (non-existing) archaic data. But we may find relevant comparisons by analyzing present-day northern dialects, which should most closely resemble the archaic forms.

4. Conclusions

The G&S results have placed the issue of Proto-Basque reconstruction in a win-win situation: all of the possible scenarios that lead to these results are interesting. The possibilities that we have considered are summarized (in reverse order) in (21):

(21) a. Proto-Basque was underlyingly the same as modern Basque (modulo the further issue of whether it independently changed for ergativity reasons).
b. Proto-Basque was underlyingly V1, and it remained thus for a very long period of time, until it lost morphological strength at C and it became SVO (or SOV, modulo ergativity considerations); the scenario presupposes a Proto-Proto-Basque that was not V1 (or did not have relevant periphrastic constructions to reanalyze).
c. The changes from Proto-Proto-Basque to Proto-Basque, and from Proto-Basque to Basque, were accelerated by drastic outside forces.

Scenario (21a) is arguably the null hypothesis, pending confirmation about how much data of the V1 sort (without the language being underlyingly V1) is required for the G&S reanalysis to occur. It is captivating because it suggests that Basque hasn't undergone significant syntactic changes (at least in this (set of) parameter(s)) for the longest time. Scenario (21b) requires an even longer stretch of time to be in play, since it involves two Core parametric re-settings, which given the hierarchy in (14) must have gone through the usual slope of Peripheral drift, Sub-set reanalysis and final Core change. It wouldn't be surprising if the necessary changes for the plausibility of scenario in (21b), which refers to a state of Proto-Basque that is prior to contact with Latin, throw us back another couple of thousand years, to the eve of civilization. Finally, scenario (21c) is perhaps the most fascinating, especially in light of claims that Proto-Basque may have had a European extension that is vaster than commonly assumed. The issue here is that this scenario requires language contact —of a pretty drastic sort. Historical times have shown relatively little structural contact between Basque and its various neighbors, but what happened in (even recent) prehistoric times is anybody's guess. It is conceivable that a broader (Proto-)Proto-Basque did have contacts with several neighbors. This is the most speculative scenario, but not because of that the least intriguing. Which adds up to showing that taking the linguistic, learnability and performance lenses seriously leads to exciting reconstructive processes, and perhaps even reliable ones.

References

Anderson, S., 1992, *A-morphous Morphology*. Cambridge: Cambridge U.P.
Baker, M., 1988, *Incorporation: A Theory of Grammatical Function Changing*. Chicago: Chicago U.P.
Bickerton, D., 1990, *Language and Species*, Chicago: Chicago U.P.
Chomsky, N., 1965, *Aspects of the Theory of Syntax*, Cambridge: MIT Press.
—, 1981, *Lectures on Government and Binding*, Dordrecth: Foris.
—, 1995, *The Minimalist Program*, Cambridge: MIT Press.
Clark, R. and I. Roberts, 1993, 'A computational model of language learnability and language change', *LI* 24, 299-345.
Crain, S. and R. Thornton, 1998, *Investigations in Universal Grammar: A Guide to Experiments in the Acquisition of Syntax and Semantics*. Cambridge: MIT Press.
Di Sciullo, A. and E. Williams, 1988, *On the definition of word*. Cambridge: MIT Press.

Dresher, E., 1998, 'Child Phonology, Learnability and Phonological Theory', in T. Bathia and W. Ritchie (eds.), *Handbook of Language Acquisition.* New York: Academic Press, 299-346.

Givon, T., 1973, 'Historical Syntax and Synchronic Morphology: An Archeologist's Field Trip', *Chicago Linguistics Society* 7, 394-425.

Gomez, R. and K. Sainz, 1995, 'On the Origin of the Finite Forms of the Basque Verb', in J. Hualde, J. Lakarra and L. Trask (eds.), *Towards a History of the Basque Language*, Current Issues in Linguistic Theory 131. Amsterdam: John Benjamins, 235-54.

Kayne, R., 1994, *The Antisymmetry of Syntax*. Cambridge: MIT Press.

Kroch, A., 2001, 'Syntactic Change', in M. Baltin and C. Collins (eds.), *The Handbook of Contemporary Syntactic Theory.* Malden: Blackwell, 699-729.

Laka, I., 1990, *Negation in Syntax*, MIT doctoral thesis.

Lighftooft, D., 1991, *How to Set Parameters: Arguments from Language Change.* Cambridge: MIT Press.

—, 1999, *The Development of Language*, Blackwell-Maryland Distinguished Lecture Series on Language and Cognition 1. Malden: Blackwell.

Rizzi, L., 1990, *Relativized Minimality.* Cambridge: MIT Press.

Trask, L., 1977, 'Historical Syntax and Basque Verbal Morphology: Two Hypotheses', in W. Douglass, R. Etulain and W. Jacobsen (eds.), *Anglo-American Contributions to Basque Studies. Essays in Honor of Jon Bilbao*. Reno: University of Nevada, 203-17.

—, 1979, 'On the origin of ergativity', in *Ergativity; Towards a Theory of Grammatical Relations*, F. Plank (ed.), London: Academic Press: 385-404.