Why Y: on the centrality of syntax
in the architecture of grammar[*]

Aritz Irurtzun

University of the Basque Country (UPV/EHU) & HiTT.
Linguistics and Basque Studies Department
Unibertsitateen Ibilbidea 5, 01006 Vitoria-Gasteiz
aritz.irurtzun@ehu.es

Abstract

I defend the centrality of the syntactic module within the general architecture of grammar. According to this model of the language faculty, syntax creates structures that are interpreted at the interfaces with the Articularory-Perceptual and Conceptual-Intentional systems. Thus, I show that the classic inverted-Y model of the architecture of grammar is better suited than alternative "parallel architectures" (*cf.* Jackendoff (1997 *et seq.*)) when accounting for interface phenomena. In order to do that, I discuss an interface phenomenon like focus that, according to some scholars, shows the need of a more articulated architecture of the grammar than the classic Y-model. I will argue that the properties of focus bear testimony to the fact that syntax outranks both interpretive modules.

Key words: architecture of grammar, Y-model, interfaces, focus, nuclear stress rule.

Table of Contents

1. Introduction

Generative grammar takes language to be a property of the mind/brain of an individual, an attained state of certain knowledge. As a mental object its design is constrained in various ways. Thus, the minimalist program tries to provide an answer to one crucial question, which I borrow from Chomsky (2005: 9-10): To what extent does language approximate an optimal solution to conditions that it must satisfy to be usable at all, given extralinguistic structural architecture?

The goal of the Minimalist Program is to go *beyond explanatory adequacy* and discover the core nature of language by pursuing two strategies (i) minimizing the assumptions and technical machinery employed in grammatical description (*i.e.* searching for the most economical and elegant explanation), but also (ii) analyzing how optimal the design of the faculty of language itself is, given the external restrictions that it has to face (the so-called *bare output conditions* or *legibility conditions*).

The aristotelian conception of language portrayed it as a vehicle to relate sound and meaning. This is a truism; the faculty of language interfaces with at least these two systems that human beings have independently of language. These two systems are called the Articulatory-Perceptual system (A-P) (sometimes also called Articulatory-Motor), and the Conceptual-Intentional system (C-I) (the thought). So our conception of the architecture of grammar should depict language as interfacing with those two systems, at least. What else is needed? Chomsky (1995 *et seq.*) argues that not much; just access to a lexicon, a repository where linguistic units (lexical items) will be stored. Those three interfaces (A-P, C-I, and lexicon) are virtual conceptual necessities without which no theory of language can expect any explanatory success.

Within those limits, the general trend in generative grammar is to consider the language faculty as a computational system that takes lexical items and *via* an operation called 'Merge' generates expressions that are interpretable at both interface levels. Conditions imposed by the language-external systems will dictate which types of expressions are 'readable' as Phonological Form (PF), the linguistic representation that interfaces with the A-P system, and Logical Form (LF), the representation that interfaces with the C-I system. Then, given that there is no apparent interaction between the C-I and A-P systems, the architecture of grammar is generally taken to have the shape of an inverted-Y, where the syntactic module generates pairs of expressions that will be interpreted in the independent A-P and C-I systems[1].

---

1. If I am not wrong, the first formulation of such a model is due to Chomsky & Lasnik (1977) where, even if it is not depicted as such, we have a description of the system (*cf.* Chomsky & Lasnik (1977: 431)). Here PF is the equivalent to the 'Universal Phonetics' (UP) of that work.

(1). *Architecture of grammar* (Y-model):

Lexicon

Language

[Phonological Form]   [*logical form*]

A-P System   C-I System
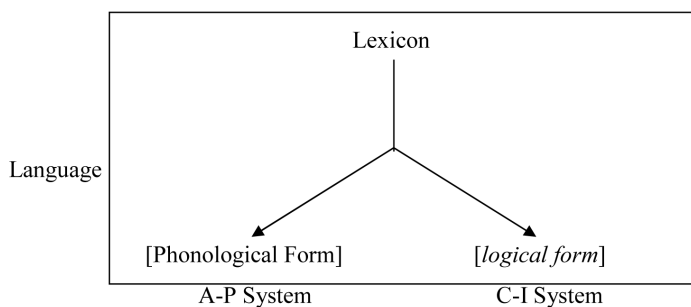
Thus, the lack of a direct link relating the PF representation and the LF representation is meant to capture the fact that a semantic feature does not affect phonological representations (and *vice-versa*)[2]. However, there are some interface phenomena that at first sight seem to be unapproachable within this model of the architecture of grammar and some scholars have proposed to modify this model and adopt a more powerful one in order to account for such a variety of phenomena as the distribution of pre and postnominal adjectives in languages like French, the creation of phonological phrases, the phenomena of cliticization, the relationship between focus, word order and nuclear stress placement and the relationship between syntax and semantics in argument structure[3]. For instance, one of the most cited piece of data that is used to argue against the classical Y-model of the architecture of grammar is the apparent mismatch between syntax and phonology in the paradigm in (2) (*cf. i.a.*, Chomsky & Halle (1986), Selkirk (1986) and Jackendoff (2002)):

(2) *Syntax*:   [$_{NP}$ this] [$_{VP}$ is [$_{NP}$ the cat [$_{CP}$ that [$_{VP}$ caught [$_{NP}$ the rat [$_{CP}$ that [$_{VP}$ stole [$_{NP}$ the cheese]]]]]]]]]

Phonology: [$_{IntP}$ this is the cat][$_{IntP}$ that caught the rat][$_{IntP}$ that stole the cheese]

2. A recent development of these ideas has lead to the constitution of the framework called 'Distributed Morphology' (DM) (*cf.* Halle & Marantz (1993) *et seq.*). The main ideas underlying the DM framework are that syntactic operations are not affected by phonological (vocabulary) nor semantic (encyclopedic) information and that all lexical items are phrasal in essence (they are composed from smaller syntactic units which are not lexical items themselves). Thus, the idea is that we have to distinguish among the syntactic items (roots and functional elements), the vocabulary (which are arbitrary relations of phonological and syntactic material) and the encyclopedia (the arbitrary relations of association of sound to meaning). DM proposes that vocabulary insertion is done at PF, and then all the derivation is reanalyzed in LF so that LF has access to which vocabulary item was inserted in each node adding to it the encyclopedic knowledge (say, whether in a given node we inserted *cat* or *dog*) (see, *e.g.*, Marantz (1997) for discussion)).
3. In fact, this type of argumentation is very popular in the literature with a functionalist orientation. See also Culicover & Jackendoff (2005) for related arguments.

The lack of isomorphism between syntactic structure and phonological structure has been used by authors like Jackendoff to claim that the conception of the architecture of grammar as being 'syntactocentric' as in (1) is just wrong and he has proposed a most radical change of the model of the grammar, postulating a "Parallel Architecture" of the grammar where all modules (phonology, syntax and conceptual structure) create their own derivations in a parallel fashion (see Jackendoff (1997, *et seq.*))[4]. Then, he proposes that there will be some correspondence rules linking the different derivations (some rules linking phonological structures and syntactic structures, some rules linking syntactic structures and conceptual structures and also some rules linking phonological structures and conceptual structures, as depicted in (3))[5]:

(3)   *Architecture of grammar* (parallel model):
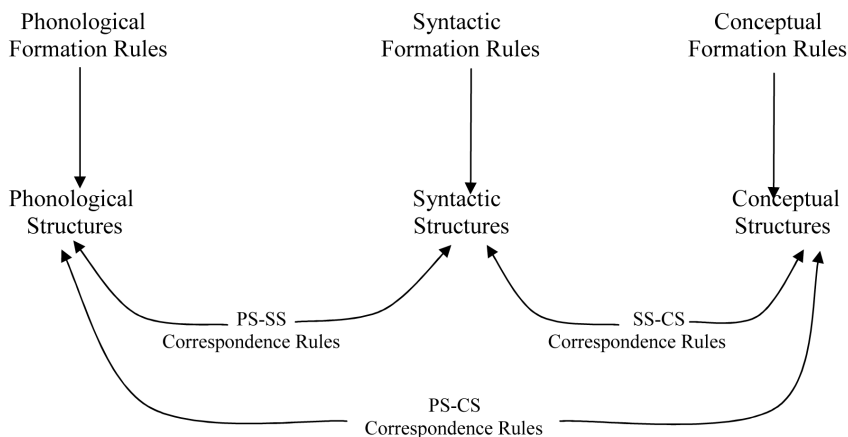
Jackendoff's proposal has been extremely successful, not only among linguists like Szendröi (2003), Camacho (2005), Dominguez (2005) or van der Hulst (2006), but also among neurologists and cognitive scientists (*cf., i.a.* Garrod & Pickering (2003), Piñango (2006))). The success of such a proposal comes with no surprise given that a more powerful system allows for more types of interrelations than a more restrictive one and phenomena that seem unapproachable within the Y-model might seem

---

4.   Actually, Jackendoff's conception of the conceptual structure is closer to what Chomsky calls the C-I system than to Chomsky's LF. In fact, according to Jackendoff "conceptual structure is not part of language per se –it is part of thought. It is the locus for the understanding of linguistic utterances in context, incorporating pragmatic considerations and "word knowledge"; it is the cognitive structure in terms of which reasoning and planning take place. That is, the hypotesized level of conceptual structure is intended as a theoretical counterpart of what common sense calls "meaning"" (Jackendoff (2002: 123)).

5.   This powerful schema allows him to capture fishy notions like the *Sense Unit Condition on Intonational Phrasing* of Selkirk (1984), a condition requiring that "the immediate constituents of an intonational phrase must together form a sense unit" (*cf.* Selkirk (1984: 286 & ff.)).

easier to capture within the more powerful architecture. However, in this article I want to discuss the predictions borne out by the inverted Y model of the architecture of grammar and compare them with those of alternative proposals like Jackedoff's. I will argue that notwithstanding the descriptive power brought by such powerful architectures, they have no explanatory power, and that they entail a bigger number of conceptual and empirical problems than the classical inverted-Y model of architecture. Thus, I will argue that the classical architecture whereby syntax is the central component of grammar is to be preferred over the alternative proposals.

In the next section I will discuss a phenomenon that has been used in the literature in order to argue for the need of a more powerful architecture of grammar: focus structure and its relation with word order and intonation. I will show the shortcomings of the arguments purporting that a new architecture of grammar offers a better explanation than the classic inverted-Y model. And furthermore, I will conclude that the model of the inverted-Y architecture of grammar is the only possible way of analyzing the facts.

2. Global interactions? A case study: Focus and the Nuclear Stress Placement

In this section I want to discuss a phenomenon that has been used to claim for the necessity of a different type of architecture of the language faculty: the phenomenon of focus, which relates syntax, phonology and semantics. I will argue that all of the arguments posited for an architecture where phonology or semantics affects syntax have to be rejected, and that the classical Y-model of the architecture of grammar offers the best explanation for the facts given that it comprises the simplest model.

A widespread observation is that many languages like English or Basque show the tendency to align nuclear stress with the focus of the sentence. Thus, a common assumption in the literature is that we can understand the fact that focus tends to bear the nuclear stress in several languages as deriving from a representational legibility condition on derivations that requires focus to have nuclear stress at PF. This is the line of research adopted by some referential works such as Selkirk (1995), Reinhart (1995, 2006), Neeleman & Reinhart (1998), Zubizarreta (1998), Zubizarreta & Vergnaud (2000), Szendrői (2001), Arregi (2003), Elordieta (2001), and Ishihara (2000), which adopt the original ideas of Cinque's (1993) *Nuclear Stress Rule* (NSR) (4) proposing some readjustments and refinements:

(4)   The *Nuclear Stress Rule* (Cinque (1993)):

    (i)   Interpret boundaries of syntactic constituents as metrical boundaries.

    (ii)   Locate the heads of line *N* constituents on line *N*+1.

    (iii)   Each rule applies to a maximal string containing no internal boundaries.

    (iv)   An asterisk on line *N* must correspond to an asterisk on line *N*+1.

That is, Cinque's proposal is that syntactic bracketing is turned into metrical bracketing in PF, and that the more embedded you are in syntax the more embed-

ded you will get in PF. Thus, the most embedded element in the metrical grid is the element that will get nuclear stress as illustrated in (5a-b) (ex. 83, p. 265 of Halle & Vergnaud (1987), nuclear stress in the item in bold):

(5a) Jesus preached to the people of **Judea**.

(5b)

.	.	.	*	Line 6
(.	.	.	*)	Line 5
.	(.	.	*)	Line 4
*	*	(*	*)	Line 3

   [Jesus [preached to the [people of Judea]]]

Staring out from this conception of the *NSR*, the approaches to focus which are based on the this rule of stress placement postulate that nuclear stress placement is what will dictate which F(ocus)-Structures are available.

While there are some differences in the implementation of their analyses, what all these approaches have in common is that they propose that the different F-Structures observed in the sentences in (6a-g) would be amenable to an explanation in terms of 'focus projection'; the element that bears the nuclear stress 'projects' its focal status to higher nodes that dominate it. Thus, with a single accent placement we can have different F-Structures:

(6a) Jesus preached to the people of [**Judea**]$_F$

(6b) Jesus preached to the people [of **Judea**]$_F$

(6c) Jesus preached to the [people of **Judea**]$_F$

(6d) Jesus preached to [the people of **Judea**]$_F$

(6e) Jesus preached [to the people of **Judea**]$_F$

(6f) Jesus [preached to the people of **Judea**]$_F$

(6g) [Jesus preached to the people of **Judea**]$_F$

Hence, according to these approaches there is no F-Structure as such in narrow syntax; rather, the *focus set* has just to be *inferred* from the placement of the nuclear stress in PF at a postgrammatical level of discourse. Any element containing the nuclear stress might be interpreted as focus, and hence, a single nuclear stress placement might mark different F-Structures. This idea implies as a natural consequence the notion of *focus projection*. According to this view, if nuclear stress falls on the most deeply embedded element in (6), this stress placement will be able to convey many different F-Structures (all the constituents containing the nuclear stress up to the whole sentence (*cf.* (6a-g)). The main consequence

that follows from this interrelation between phonology (*via* the nuclear stress), syntax (the properties if clausal architecture), and semantics (the focus of a sentence) is that a sentence will not have an 'actual focus' *per se* but rather 'a set of possible foci', that is, the set of nodes that an actual nuclear stress placement can mark as focused. This idea is explicitly stated, for instance, in Reinhart (2006: 158)[6]:

(7)   *The focus set*: The *focus set* of a derivation D includes all and only the constituents that contain the main stress of D.

This focus rule accounts for the basic facts regarding focus projection in the different sentences in (6). Under this approach, in a SVO sentence of a language like English the nuclear stress on the object may mark as focused either the direct object itself, the VP, or the whole clause. According to Reinhart (1995, 2006), in the unmarked case the *NSR* assigns nuclear stress to the most embedded position. The 'focus projection' is seen as automatic, and then the discourse will decide what the actual focus is from among the elements in the *focus set*.

However, as observed by Reinhart (1995, 2006), this strategy will not serve to mark focus on certain elements like the subject or the verb since, clearly, they do not contain the nuclear stressed element (providing that, as said before, by default the *NSR* assigns the nuclear stress to the most deeply embedded position in absolute terms (*i.e.*, the direct object)). Thus, according to Reinhart (1995, 2006), in order to mark focus on a phrase that cannot be focused by the default focus projection of the object (*i.e.*, a phrase which is not in the original *focus set*), some marked strategies must be employed. For instance, in English-like languages —where focus does not affect word order—, a deaccentuation rule will deaccent the object and a marked stress rule will assign nuclear stress to whichever element has to be interpreted as focused (8)[7]:

(8)   *Accent Shift*:

Step$_1$ (base) ⇨ SVO

Step$_2$ (*NSR*) ⇨ SV<u>O</u> (*focus set:* {O, VP, TP, CP})

Step$_3$ (deaccentuation) ⇨ SVO

Step$_4$ (marked stress placement) ⇨ <u>S</u>VO (*focus set:* {S, TP, CP})

On the other hand, there are also discourse-configurational languages like Hungarian or Basque, where focus affects the basic order of constituents. See for

---

6.   Selkirk (1995) proposes an analogous rule of focus projection from heads to phrases that has also been very influential (*cf. i.a.* Schwarzschild (1999)). The main difference between the proposals in Selkirk (1995) and Reinhart (1995, 2006) is that for the former the PF rule of projection applies optionally and for the latter it applies automatically (creating focus sets).

7.   Here I mark nuclear stress with boldface and with an underline for clarity.

example the data in (9a) for an *out-of-the-blue* utterance (sentence-focus) and (9b) for a subject-focus sentence in Basque:

(9a) [Ainarak **ardoa** erosi zuen]$_F$.
     Ainara   wine buy AUX
     "Ainara bought wine"

(9b) Ardoa [**Ainarak**]$_F$ erosi zuen.
     wine Ainara   buy AUX
     "Ainara bought wine"

   It is posited that those elements that are most deeply embedded must scramble to a position higher up in the structure (see Arregi (2003) for this type of proposal). As a result of this scrambling operation, the element that has to be interpreted as focus becomes the most deeply embedded one and thus it can receive nuclear stress as roughly illustrated in (10):

(10) *Displacement*:

   Step$_1$ (base) ⇨ SOV

   Step$_2$ (*NSR*) ⇨ SO̲V (*focus set:* {O, VP, TP, CP})

   Step$_3$ (deaccentuation) ⇨ SOV

   Step$_4$ (*scrambling of O over S*) ⇨ OSV

   Step$_5$ (*NSR*) ⇨ OS̲V (*focus set:* {S, TP, CP})

   Having the nuclear stress on the subject, a new *focus set* is created and the elements in this set (in particular the subject) can be interpreted as focus (see below).
   So, these two strategies (namely, (i) stress shifts in languages like English, and (ii) scrambling of unfocused constituents in languages like Basque) provide the intended focus-nuclear stress correlation at PF.[8]
   There is still a question for this type of approach, though. According to the focus rule in (6), nuclear stress placement on a subject should also be able to denote sentence-focus. This is so because the CP, dominating all material, contains the nuclear stressed item when it is on the subject. Likewise for TP[9]. Thus, if we apply any of the marked operations proposed by *NSR*-based approaches that I have illustrated in (8) and (10), then, the *focus set* of a sentence with nuclear stress on the subject is {Subject, TP, CP} as illustrated in (10). However, Reinhart (1995, 2006) argues that this is not an available option because it is *anti-economical*: the sen-

---

8. *Cf.* Reinhart (2006) for the specificities of these operations.
9. Let me assume for the sake of the argument that the subject is in Spec-TP, even though Arregi (2003) assumes that it sits in Spec-vP and that the verb is higher up (that would be the reason for the subject to receive the nuclear stress instead of the verb).

tence-focus interpretation could have been obtained *via* the regular projection algorithm from the nuclear stress on the object in the base configuration (just with the steps 1 and 2 in (8) and (10)) and without having to incur into marked operations. That is, Reinhart (1995, 2006) argues that it is more economical to interpret sentence focus from the projection of the *NSR* that would apply to the object by default than to interpret it as the outcome of three operations: (i) default *NSR* to the object first, (ii) deaccenting it, and (iii) applying the 'marked' stress shift of (8) to assign nuclear stress to the subject. Therefore, Reinhart (1995, 2006) concludes that an economy principle would prevent that choice. Under Reinhart's theory, economy is calculated by comparison of both derivational possibilities for interpreting focus on the whole clause (what Reinhart (2006) calls a 'reference-set computation'). Thus, even if both *focus set*s of the sentences in (11) and (12) (with nuclear stress on the object and subject respectively) contain TP and CP, the choice of TP or CP as the actual focus in a sentence with nuclear stress on the subject would be anti-economical, as those options would also be available with nuclear stress on the object[10]:

(11) John likes **wine**.

   Focus set: {Obj, VP, TP, CP}

(12) **John** likes wine.

   Focus set: {Subj, *TP*, *CP*}

Based on this view of the *NSR* and F-Structure, a very productive program of research has been recently developed refining the details of the theory and applying it to previously unstudied languages (*cf. i.a.*, Reinhart (1995, 2006) and Neeleman & Reinhart (1998) for English and Dutch, Zubizarreta (1998) and Dominguez (2004) for Spanish, Elordieta (2001) and Arregi (2003) for Basque, Szendröi (2003) for Hungarian, English and Italian, Ishihara (2003) for Japanese, Arnaudova (2003) for Bulgarian and Georgiafentis (2004) for Greek). I will refer to this program of research as the '*NSR*-based theories of F-Structure', given that what all these proposals have in common is that, according to them, there is no focus phrase *per se* during the syntactic computation to the interfaces. The gist of all these approaches is that there is a legibility condition on derivations that requires focus to have nuclear stress at PF and focus is calculated taking into account the phrases that contain the nuclear stress in the PF representation. Under this view, then, nuclear stress placement is prior to the setting of F-Structure. As a result, the F-Structure of a sentence is computationally ambiguous (since the accented element denotes, *via* focus projection, a set of different F-Structures). Obviously, this view entails serious conceptual problems for the Y-model of the architecture of the

10. Following Reinhart's notation, potential foci which are not available due to their 'anti-economical' nature (even though they are in the focus set) are marked with italics.

grammar, given that it is assumed that some operations in narrow syntax are taking place in order to (i) leave the element to be focused in the most embedded position (in order to get nuclear stress), and (ii) interpret from this nuclear stress placement the focus-background partition of the sentence. As a consequence, all these approaches require a radical change in our conception of the architecture of grammar, syntax and semantics have to have access to phonological representations.

In the next section I will discuss the conceptual shortcomings of this type of theory arguing that it is weaker than the classical Y-model of the architecture of grammar regarding their explanatory power.

2.1. *Shortcomings of* NSR-*based Theories of the F-Structure*

Due to space limitations I will not discuss here the varied empirical evidence against *NSR*-based theories of the F-Structure (see Irurtzun (2007, 2008) for details and discussion). Here I just want to point out that there is a cornucopia of empirical arguments against both the idea of the focus sets and the purported legibility condition requiring F-Structures to bear nuclear stress in PF. The evidence shows that a determinate F-Structure is present in the PF representation (which can be observed in the appearance of specific pitch accents and boundary tones marking the actual F-Structure), and the fact that many languages do not have any focal accent also shows that we cannot say that there is a legibility condition (by definition, universal) requiring F-Structures to bear nuclear stress. Furthermore, there are many constructions where the focal element that bears the nuclear stress is not in the absolutely most embedded position, which makes the initial axiom of NSR-based theories of focus more than dubious.

Here I want to discuss some of the conceptual drawbacks of theories that purport that the derivation of a sentence is affected by its phonological structure, and I want to show what it teaches us about the architecture of grammar. I will be focusing on three main problems: (i) the lack of explicitness of the *NSR*-based approach to F-Structure, (ii) the problem of the setting of F-Structure, and (iii) the fuzzy nature of the purported *focus set*.

The first criticism I would like to make here concerns the lack of explicitness of *NSR*-based approaches. As explained before, their main premises are based on a purported interface legibility condition that requires focus to have the nuclear stress at PF. My goal is to show that this argumentation is paradoxical in its conception of F-Structure: the crucial idea of *NSR*-based approaches is that there is no F-Structure *per se* during the computation from the Numeration to PF, the actual focus is determined *after* nuclear stress placement. But on the other hand the system needs to know what the focus will be in order to trigger the operations that lead to assign nuclear stress within it. That is, there is a paradox concerning where focus appears: on the one hand, *NSR*-based theories involve the assumption that F-Structure is somehow there at the beginning of the derivation, that is, the *NSR* applies either regularly or through a marked operation (stress shift/scrambling), depending on the intended F-Structure. But on the other hand, it is claimed that PF is responsible for determining the possible F-Structures (the *focus set*) given the *NSR* and focus pro-

jection. Basically, the question is the following one: if there is no F-Structure from the beginning, then *ceteris paribus* the *NSR* should invariably assign nuclear stress to the most deeply embedded element, and there would be no reason for the scrambling movements of languages like Basque, nor for the marked stress assignments of languages like English. On the other hand, if the F-Structure is there from the beginning, then there is no need to postulate ambiguities, *focus sets*, nor focus projections; it is just there. As I will discuss next, the problem is that the argumentation of *NSR*-based theories of F-Structure is necessarily circular: the focal XP of a sentence will be inferred from the placement of the nuclear stress, but on the other hand, the position of the nuclear stress will depend on what the actual 'semantic' focus is. Clearly, this type of argumentation is circular.

Furthermore, if we assume a Y-model of the architecture of the grammar, then, a condition imposed on PF will be unable to drive narrow syntactic operations, that is, there is no *look ahead*. And, unless specified further, the displacements that the *NSR*-based theories of F-Structure postulate are essentially *look ahead* operations. Obviously this issue becomes problematic provided that we accept the classical Y-model of the architecture of the grammar. And in fact, as I said before, many of the theorists adopting *NSR*-based theories of F-Structure postulate that the conclusion of their work is that the architecture of the grammar has to be changed radically in order to accommodate for focus phenomena. Thus, for instance, Zubizarreta (1998) proposes a novel architecture of the grammar along the lines in (13). According to this model, the derivation unfolds creating sets of phrase markers until one single phrase marker is obtained (Σ-Structure). Then some operations like F-marking, *NSR* and prosodic movements take place until we reach the level of LF. At this stage the derivation branches in two branches, one that derives in a PF representation and the other one that provides what Zubizarreta (1998) calls the "Assertion Structure", the information structure of the sentence where the focus-presupposition partition is encoded[11]:

(13)

Lexicon

Σ-Structure

LF

PF   Assertion Structure

---
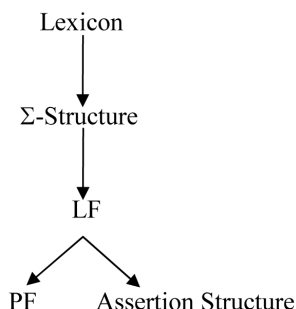
11. *Cf.* also Vallduví's (1995) architecture with a similar level of "Information Structure", and Erteschik-Shir & Strahov's (2004) "I-model".

Zubizarreta's (1998) model assumes that all the focally relevant operations of *NSR* and "prosodic movements" take place in the stretch between Σ-Structure and LF. However, according to the architecture depicted in (13) at this stage there is still no phonological representation, so even changing the architecture of the grammar all these operations take place *looking ahead* (that is, just in order to get the intended representation at a later derivational point).

Furthermore, it is assumed that the final phonological representation derives from the LF representation (which, I assume, provides also the interface with the C-I system). But then, what type of variability do we expect in PF if its input is the output of the logically unambiguous and universal representation in LF? If we adopt the architecture in (13), the LF→PF stretch should be very powerful in order to allow for all the variability we observe in the surface of natural languages. Basically, the LF→PF stretch would be equivalent to the D-Structure→PF strecht of the Principles and Parameters model, so instead of simplifying the grammar we would complicate it exponentially (a syntactic derivation would take place until Σ, from Σ to LF the focally relevant operations would take place deriving in an unambiguous LF representation and then the PF representation would be obtained departing from this LF representation). That is, all word order, phononological phrasing and morphological differences (among others) that we observe in natural languages would be just idiosyncratic properties of languages (non-traceable to their syntax *stricto sensu*). Let me illustrate this: imagine that we want to account for the derivation of the subject-focus sentences in (14a-b) (English and Basque, respectively):

(14a) [**Ainara**]$_F$ bought wine.

(14b) [**Ainarak**]$_F$ erosi zuen   ardoa.

      Ainara   buy   AUX wine

      "Ainara bought wine"

A proper logical form representation of these sentences (leaving aside the contribution of focus) could be the one in (15), which says, roughly, that there is a past event of buying whose theme is wine and whose agent is Ainara[12]:

(15) ∃e[Buying(e) & Past(e) & Theme(e, wine) & Agent(e, Ainara)]

If we adopt Zubizarreta's (1998) architecture, the derivation would unfold for both languages in a similar fashion up to the point of LF, where we have the representation in (15). Now, in order to account for the surface differences in English and Basque (or in any other language), we would have to posit a variety of operations taking place in the stretch from LF to PF, operations that would be phono-

---

12. Here I am using a Neo-Davidsonian notation but the argument could equally be stated using other types of representations.

logical in essence and that would have no relation with semantics. That is, all the regularities and differences found in natural languages would be due to purely phonological operations. As I said before this view of the architecture of grammar complicates the LF-PF stretch exponentially.

On the other hand, some authors like Jackendoff have proposed still more powerful "parallel architectures" in order to allow phonology and semantics to affect syntax (and each other). These parallel architectures where any module is allowed to affect any other are much more debated, and they are often invoked whenever an interface phenomenon escapes from a straightforward explanation, especially in the syntax-phonology interface (*cf. i.a.* Szendröi (2001), Dominguez (2005))[13]. In fact, one of the most popular arguments used for arguing in favor of the jackendoffian parallel architecture is the fact that focus constructions display phonological, syntactic, semantic and/or morphological effects (*cf.* the references cited above and specially Szendröi (2001) and Jackendoff (2002)). This issue deserves a closer inspection: as I said before, in the jackendoffian parallel arquitecture "correspondence rules" are introduced to link the independent derivations. However, I would like to argue that this model as well incurs in a global *look ahead*. The purported derivational operations that take place to provide the intended representations are unmotivated otherwise; they are posited in order to arrive at PF representation. However, the system is supposed to be derivational in essence, there is nothing like 'move α', so no way to obtain the intended representations. It is, in my view, like an Optimality Theoretic grammar but without the generator function to provide the infinite set of candidates to be evaluated[14].

But more importantly, the idea of independent derivations is extremely dubious: what type of phonological representation are we going to build independently of syntax? And how are they going to affect syntactic structure building? For instance, given a syntactically unordered set of words in a simplified numeration like (16), how does the computational system build phonological structures if it does not take as input the output of syntax?

(16)   NUMERATION (simplified): {these, words, are, not, syntactically, combined}

The traditional view on phonological phrasing is one in which there are some rules whose structural description is a syntactic structure and whose structural change provides a phonological phrase of some sort (*i.e.*, a syntax-to-phonology mapping). In fact, in the pre-OT literature two were the main proposals for phonological phrasing algorithms: Selkirk's (1986) and Nespor & Vogel's (1983). The difference among them was that Selkirk's algorithm proposed an end-based phonological phrasing (17) whereas Nespor & Vogel's proposed a relation-based one (18):

---

13. See also van der Hulst (2006) for an interesting discussion on the place and nature of the phonological component.
14. See Szendröi (2001) for a proposal mixing Brody's (1995) representational Lexico-Logical Form, a jackendoffian architecture of parallel derivations, and an optimality theoretic grammar for the interfaces.

(17)   The idea is that a unit of phonological structure, a derived domain, will have
       as its terminal string the stretch of the surface syntactic structure that is
       demarcated by the right or left ends of selected constituents, as in (i)

   (i)   a. $_\alpha$[…, where ... contains no $_\alpha$[

         b.   …]$_\alpha$, where ... contains no]$_\alpha$

       The general claim I am making here is that α will be drawn from the set of
       categories defined in the X-bar theory, and that α indicates only a level (or
       type) in the X-bar hierarchy, which is to say that the syntax-to-prosodic struc-
       ture mapping is claimed to be cross-categorial in nature.

                                                                    (from Selkirk (1986: 385))

(18) φ construction: join into a φ any lexical head X with all items in its nonre-
       cursive side within the maximal projection and with any other nonlexical
       items on the same side (e.g. prepositions, complementizers, conjunctions,
       copulas).
                                                            (from Nespor & Vogel 1983: 124)

       With the advent of OT, constraint-based phonological phrasing algorithms have
been proposed, but even in these algorithms, a syntactic structure is taken to be the
input to the harmony computation. An example of this way of thought is
Truckenbrodt's WRAP-XP constraint, which makes reference to the XPs which
were built up in the syntactic component:

(19)   WRAP-XP

       Each XP is contained in a phonological phrase.

                                                            (from Truckenbrodt (1999: 228))

       However, Jackendoff's type of parallel architecture postulates an independent
phonological derivation, which does not take syntactic structure as input. Note that
crucially, if it did, then it would be syntactocentric in essence (as Jackendoff some-
how recognizes (*cf.* Jackendoff 1998: 23)). As we advanced before, Jackendoff
(1997: 28-29) postulates that the syntax-phonology interface is but a list of corre-
spondence rules between independently built syntactic phrases and phonological
phrases, but we are never told how to build a phonological phrase independent of
syntax. All the jackendoffian correspondence rules do for us is to provide relations
like the one in (20):

(20)   If a syntactic constituent X_1 corresponds to a phonological constituent Y_1,
       and a syntactic constituent X_2 corresponds to a phonological constituent Y_2,
       then the linear order of X_1 and X_2 preferably corresponds to the linear order
       of Y_1 and Y_2.
                                                            (from Jackendoff (1997: 28))

That is, Jackendoff denies any mapping *operation* from syntax to phonology; there are just *correspondences* between two independent representations. The question is: how are those representations obtained? A customary assumption is that the processes of syntactic structure building and linear ordering are obtained from selection and c-command (*via* Kayne's (1994) *Linear Correspondency Axiom*[15]). But how does phonology work to create structure from the lexical items? Is there any sort of 'Merge' operation in the phonological component? How do we get the phonological representations that are to be corresponded to the syntactic representations? Jackendoff's system does not offer an answer. In fact, whenever Jackendoff presents the syntax-phonology correspondence, he starts out from already built structures like the previously cited (21) and then formulates the optimal correspondence rules like (22), where 'IntP' stands for Intonational Phrase (*cf.* Jackendoff 2002: 119-120):

(21)
> *Syntax*: $[_{NP}$ this$]$ $[_{VP}$ is $[_{NP}$ the cat $[_{CP}$ that $[_{VP}$ caught $[_{NP}$ the rat $[_{CP}$ that $[_{VP}$ stole $[_{NP}$ the cheese$]]]]]]]]]$
>
> *Phonology*: $[_{IntP}$ this is the cat$][_{IntP}$ that caught the rat$][_{IntP}$ that stole the cheese$]$

(22)   a. An utterance consists of a series of one or more concatenated IntPs forming a flat structure. Each IntP is a sequence of Words.

     b. *Preferably*, the IntPs are of equal length.

     c. *Preferably*, the longest IntP is at the end.

     d. (Possibly, some strong preferences on maximum duration of IntPs, e.g. try not to go over three seconds).

The interface constraint relating IntPs to syntactic structure would be roughly as stated in (23):

(23)   An IntP corresponds to all of a syntactic constituent C, except that a sub-constituent at the right-hand end of C can be omitted.

Thus, according to Jackendoff, the mismatch of (21) is explained as follows (*cf.* Jackendoff 2002: 120): "The first IntP corresponds to the whole sentence, but the large relative clause *that caught the rat that stole the cheese* is omitted. In turn, this relative clause is further carved: the final relative clause is taken out to form the third IntP. The result is perfectly balanced prosodically."

The theory sketched describes the facts nicely. But as can be observed, there is no syntax-independent process of phonological phrase construction. On the contrary, the process in (22) is defined as applying upon the different syntactic cycles.

---

15. See also Fortuny & Corominas (this volume).

This implies that the syntactic representation precedes in derivational terms the phonological phrase construction. That is, there were no phonological phrases before, for the number of IntPs depends directly on the number of CPs. So, in the end, the alleged correspondence rule is more like the classic syntax-phonology mapping. In other words, there are some *operations* that take as input syntactic phrases and give as output phonological phrases; however complex this process might be[16,17]. So, it seems that here as well, there is *no escape from syntax*, and the rules of phonological phrasing and nuclear stress placement have to take as their input a syntactic structure.

The question, then, is how to account for the phonological and syntactic properties of focus. The goal of this article was not to address this specific issue as such, but to show that any other attempt for an alternative architecture of grammar is paradoxical and/or vacuous. For the focal facts, in Irurtzun (2007, 2008) I have proposed a derivational analysis of F-Structure building whereby the F-Structure is built-up in the syntactic component and strictly adhering to the classical Y-model of the architecture of grammar. This allows the F-Structure to be the subject of the syntactic (A'-like) movements which are described in the cartographic literature (*cf. i.a.*, Rizzi (1997)) and furthermore, by being set in the narrow syntax, it can also take place in the structural description of purely phonological operations like nuclear stress assignment, which would take place within the F-Structure:

(23) **Focal NSR (FNSR):** Assign nuclear stress to the most embedded element *within the F-Structure*.

This rule, along a rule assigning a specific prosodic phrase to the F-Structure, accounts for the PF effects of focalization in a natural way and furthermore, by having the F-Structure built up in the syntactic component, allows it to have a semantic interpretation without incurring in intermodular operations of the PF-LF sort. Syntax construes and the interfaces interpret.

3. Summary and conclusions

I have argued that syntax outranks both semantics and phonology and hence, that the classical 'syntactocentric' architecture of grammar is the best model we have.

---

16. Obviously, syntactic phrases are not *isomorphic* to phonological phrases (which, in no case proofs the mistake of syntactocentrism). There will always be some readjustments due to phonological reasons like length, accent placements and so on (that is precisely the role of Jackendoff's "preferably" clause in (22)). The fact that those mismatches are due to readjustements (in derivational prosodic phrasing algorithms) or unfaithful mappings (in OT grammars) shows us that syntactocentrism is on the right track.

17. Note that, furthermore, in various syntactocentric models phonology even can access the output of syntax several times in the derivation, as in Uriagereka's (1999) *Multiple Spell Out* or Chomsky's (2000 *et seq.*) *phase*-based models (*cf.* Dobashi (2003) and Samuels (2009) for a much detailed analysis of relevant data). More specifically, "syntactocentric" explanations have been proposed for this specific piece of data, as in Wagner (2007).

Alternative conceptions do not help understanding the phenomena for they necessarily incur in global *look ahead* operations and paradoxes. I have shown that more powerful architectures like the parallel architecture of Jackendoff are unable to account for the facts of phonological phrase construction and nuclear stress placement. If we want to even describe the data we need to conceive the grammar as syntactocentric in essence, whereby semantic or phonological notions are unable to drive syntactic operations. In fact, in my opinion the type of evidence that would show that phonology does affect syntax and hence that syntactocentrism is wrong would be to observe the application of a purely phonological rule triggering the application of a purely syntactic operation. For instance something like the invented rule depicted in (24):

(24)   [+ANT, COR] → [-ANT, COR] / [+HIGH, -ROUND]+__X]$_v$] → V1

This is a contextually restricted rule of palatalization of a dental stop within a verb (the structural description), triggering a syntactic rule of verb fronting (the structural change): A hypothetical paradigm corresponding to this sort of rule would be the "fake-Basque" of (25) (a completely made-out paradigm which does not exist in Basque, it is just to illustrate my point):
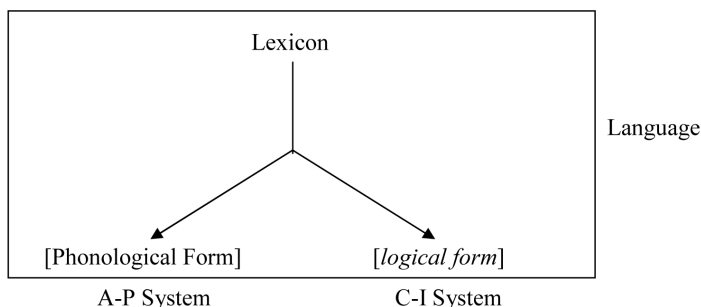
(25a) *Base*: Jonek zopa lodi[affix] du   ⇨   SOV
            Jon   soup   thicken   AUX

(25b) Jonek zopa loditzen du   ⇨   SOV
      Jon   soup thicken AUX
      "Jon thickens the soup"

(25c) *Loditzen Jonek zopa   du   ⇨   *VSO
      thicken   Jon   soup AUX
      "Jon thickens the soup"

(25d) *Jonek zopa lodittu   du   ⇨   *SOV
      Jon   soup thicken AUX
      "Jon thickened the soup"

(25e) Lodittu Jonek zopa du   ⇨   VSO
      thicken Jon   soup AUX
      "Jon thickened the soup"

So, imagine that Basque had a pattern like (25), where depending on the aspectual affix that the verb 'lodi' (to thicken) takes, the syntax of an *out-of-the-blue* sentence changes. Imagine that if it takes the suffix '-tzen' (punctual/habitual, where 'tz' is a dorsal affricate) no phonological rule applies and the syntax does not change

(the rule in (24) does not apply). Thus we have the regular SOV word order we saw so far (*cf.* (25b-c)). But imagine that it takes the suffix '-tu' (perfective). That would trigger a phonological palatalization (T→C) and, as a consequence rule (24) applies raising the verb and ending in a VSO order (25d-e).

This type of evidence is something that, up to my knowledge, has not been attested in any language. I believe that this indicates that the classical "syntactocentric" model is essentially correct and syntax is "phonology-free" (*cf.* Miller, Pullum & Zwicky (1997)). Syntax creates structure that is then interpreted by the interfaces:

(26) The Architecture of Language (Y-model)

```
                    Lexicon

                                          Language


    [Phonological Form]        [logical form]
      A-P System                 C-I System
```

Thus, given that the types of possible relations among the different grammatical modules are encoded in the structure of the architecture of the language faculty itself (unlike in Jackendoff's model, which are just language-specific rules) we can expect a clear boundary on the type of interactions that we will not see in natural languages. And thus, data like (25) are predicted not to exist in any natural language.

In more general terms, we can conclude with a positive answer to the minimalist question; that is, that the core nature of language presents an optimal design as a bridge in between the interface components and given the external restrictions that it has to face (the so-called *bare output conditions* or *legibility conditions*).

References

Arnaudova, Olga (2003). *Focus and Bulgarian clause structure*, University of Ottawa, doctoral dissertation.

Arregi, Karlos (2003). *Focus on Basque Movements*, MIT, doctoral dissertation.

Camacho, María Victoria (2005). "En torno a la arquitectura de la gramática en el modelo generativista: revisión y nuevas perspectivas". In Nepomuceno, Ángel, Salguero Francisco José; Soler, Fernando (eds.). *Estudios de Lógica y Lenguaje* I. Seville: Mergablum, pp.125-146.

Chomsky, Noam (2005). "Three factors in language design". *Linguistic Inquiry* 36: 1-22.

Chomsky, Noam (2000). "Minimalist inquiries: the framework". In Martin, Roger. Michaels, David; Uriagereka, Juan (eds.). *Step by step: essays on minimalist syntax in honor of Howard Lasnik.* Cambridge: MIT Press, pp. 89-155.

Chomsky, Noam (1995). *The minimalist program,* Cambridge: MIT Press.

Chomsky, Noam; Lasnik, Howard (1977). "Filters and Control". *Linguistic Inquiry* 8-3: 425-504.

Cinque, Guigliemo (1993). "A null theory of phrase and compound stress". *Linguistic Inquiry* 24: 239-298.

Culicover, Peter W.; Jackendoff, Ray (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Dobashi, Yoshihito (2003). *Phonological phrasing and syntactic derivation*. Cornell University, doctoral dissertation.

Dominey, Peter Ford (2003). "A Conceptuocentric Shift in the Characterization of Language". *Behavioral and Brain Sciences* 26-6: 674-675.

Elordieta, Arantzazu (2001). *Verb movement and constituent permutation in Basque.* Leiden University, doctoral dissertation.

Erteschik-Shir, Nomi; Strahov, Natalia (2004). "Focus Structure Architecture and P-Syntax". *Lingua* 114: 301-323.

Freidin, Robert, "Imaginary Mistakes versus Real Problems in Generative Grammar". *Behavioral and Brain Sciences* 26-6: 677-678.

Garrod, Simon; Pickering, Martin J. (2003). "Linguistics fit for dialogue". *Behavioral and Brain Sciences* 26-6: 678-678.

Georgiafentis, Michalis (2004). *Focus and word order variation in Greek*. Reading University, doctoral dissertation.

Gervain, Judith (2003). "Where is the Lexicon?". *Behavioral and Brain Sciences* 26-6: 678-679.

Halle, Morris; Marantz, Alec (1993). "Distributed Morphology and the Pieces of Inflection". In Hale, Kennet; Keyser, Samuel-Jay (eds.). *The View from Building 20*. Cambridge: Mit Press, pp. 111-176.

Halle, Morris; Vergnaud, Jean-Roger (1987). *An essay on stress*. Cambridge: MIT Press.

van der Hulst, Harry (2006). "On the Parallel Organization of Linguistic Components". *Lingua* 116: 657-688.

Irurtzun, Aritz (2008). "A Derivational Approach to the Focus Structure". *Canadian Journal of Linguistics* 53-2/3: 355-386.

Irurtzun, Aritz (2007). *The Grammar of Focus at the Interfaces*. University of the Basque Country (UPV/EHU), doctoral dissertation.

Ishihara, Sinichiro (2000). "Stress, focus, and scrambling in Japanese". *MIT Working Papers in Linguistics* 39: 142-175.

Jackendoff, Ray (2007a). "A Parallel Architecture Perspective on Language Processing". *Brain Research* 1146: 2-22.

Jackendoff, Ray (2007b). *Language, Conciousness, Culture: Essays on Mental Structure*. Cambridge: MIT Press.

Jackendoff, Ray (2003). "Toward better mutual understanding". *Behavioral and Brain Sciences* 26-6: 695-707.

Jackendoff, Ray (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jackendoff, Ray (1998). "The Architecture of the Language Faculty: A Neominimalist Perspective". In Culicover, Peter; McNally, Louise (eds.). *Syntax & Semantics: The Limits of Syntax* (vol. 29). New York: Academic Press, pp. 19-46.

Jackendoff, Ray (1997). *The Architecture of the Language Faculty*. Cambridge: MIT Press.

Kayne, Richard (1994). *The Antisymmetry of Syntax*, Cambridge: MIT Press.

Marantz, Alec (1997). "'Cat' as a phrasal idiom". MIT, unpublished manuscript.

Miller, Philip H.; Pullum, Geoffrey; Zwicky, Arnold M. (1997). "The principle of phonology-free syntax: four apparent counterexamples in French". *Journal of Linguistics* 33. 67-90.

Neeleman, Ad; Reinhart; Tanya (1998). "Scrambling and the PF interface". In Butt, Miriam; Geuder; Wilhem (eds.). *The projection of arguments; Lexical and compositional factors.* Stanford: CSLI Publications, pp. 309-353.

Piñango, María Mercedes (2006). "Understanding the architecture of language: the possible role of neurology". *Trends in Conitie Sciences* 10-2: 49-51.

Reinhart, Tanya (2006). *Interface strategies: Optimal and costly computations*. Cambridge: MIT Press.

Reinhart, Tanya (1995). "Interface strategies". *OTS Working Papers in Linguistic,* TL-95-002, Utrecht: Utrecht University.

Rizzi, Luigi (1997). "The fine structure of the left periphery". In Haegeman, Liliane (ed.). *Elements of grammar*. Dordrecht: Kluwer, pp. 281-337.

Samuels, Bridget (2009). *The Structure of Phonological Theory*, Harvard University, doctoral dissertation.

Schwarzschild, Roger (1999). "GIVENness, AVOIDF, and other constraints on the placement of accent", *Natural Language Semantics* 7: 141-177.

Selkirk, Elisabeth O. (1995). "Sentence prosody: Intonation, stress, and phrasing". In Goldsmith, John A. (ed.). *The handbook of phonological theory*. Oxford: Blackwell, pp. 550-569.

Selkirk, Elisabeth O. (1986). "On derived domains in sentence phonology", *Phonology Yearbook* 3: 371-405.

Szendrői, Kriszta (2001). *Focus and the syntax-phonology interface,* University College London, doctoral dissertation.

Truckenbrodt, Hubert (1999). "On the relation between syntactic phrases and phonological phrases". *Linguistic Inquiry* 90-2: 219-255.

Uriagereka, Juan (1999). "Multiple spell out". Epstein, Samuel David; Hornstein, Norbert (eds.). *Working minimalism*. Cambridge: MIT Press, pp. 251-282.

Vallduví, Enric (1995). "Structural properties of information packaging in Catalan". In Kiss, Katalin É. (ed.). *Discourse configurational languages*. Oxford: Oxford University Press, pp. 122-152.

Wagner. Michael (2007). "Prosody and Recursion in Coordinate Structures and Beyond". Cornell University, unpublished manuscript.

Zubizarreta, Maria Luisa (1998). *Prosody, focus and word order.* Cambridge: MIT Press.

Zubizarreta, Maria Luisa; Vergnaud, Jean-Roger (2000). "Phrasal stress and syntax". In van Oostendorp, Mark; Anagnastopoulou, Elena (eds.). *Progress in grammar: Articles at the 20th anniversary of the comparison of grammatical models group in Tilburg*, electronic publication. URL: http://www.roquade.nl/meertens/progressingrammar/index.html