Word Order Typology in Optimality Theory*

JOÃO COSTA
(Universidade de Lisboa & HIL\Leiden University)

1. Introduction.

SVO, SOV, VSO and VOS are all well-attested basic word orders in different languages. It is not clear however how to formally motivate why this type of variation at the base exists.

In this paper, I intend to look at unmarked word orders in different languages and explain why there is variation at the base. I will propose an analysis within Optimality Theory (Prince and Smolensky 1993), and suggest that the observed variation be due to the effect of constraints that are normally not active due to the domination of higher ranked discourse-related constraints. I will therefore suggest that unmarked word orders are a case of the more general phenomenon Emergence of the Unmarked (McCarthy and Prince 1994).

The organization of the paper is the following: section 2 illustrates the word order patterns to be looked at. Section 3 spells out theoretical assumptions, and section 4 the constraints to be used, and the ranking proposed for Portuguese. In section 5, I derive the other word orders considered by re-ranking the constraints proposed for Portuguese.

This work intends to test the power of OT as a theory of language variation, and to add to the studies done by Grimshaw and Samek-Lodovici (1995,1996), Samek-Lodovici (1996), Costa (1996) and Choi (1996), who suggest that different word orders are not optional, but the result of different functional specifications in the input. Here, the reverse situation is explored: if several word orders are legitimate expressions of the functional specification in the input, why and how does a language select one of them as basic?

* This paper is a preliminary version of a chapter of my dissertation (Costa (forthcoming)). A slightly modified version of this article will appear in G. Légendre, S. Vikner & J. Grimshaw (eds.). OT Syntax. MIT Press. I thank my thesis advisors Teun Hoekstra and Johan Rooryck for comments and support.

2. Problem.

A well-known fact concerning linguistic variation is that different languages display different basic word orders. By basic word order, it is here meant the word order emerging in out-of-the blue contexts, or in sentence-focus contexts. A clear test for this is to check which word order emerges as an answer to a question like *what happened?* (see Li 1976, Dik 1978, among many others).

The following examples show different base word orders for different languages:

Portuguese is an SVO language:

(1) *Portuguese:*
O João comeu o bolo.
John ate the cake

References to the basic word order in Portuguese are made in Ambar (1992), Duarte (1987), Mateus et alii (1989), among many others.

In other Romance languages, the same basic word-order can be found: as described by Rizzi (1982), among others, Italian also displays SVO as its basic word-order:

(2) *Italian:*
Andrea ha letto il libro.
Andrea has read the book

Likewise, Spanish exhibits the same word order, according to the description made in Hernanz and Brucart (1987):

(3) *Spanish:*
Juan ha visto a Maria.
Juan has seen Maria

It is worthy noting that the situation in Spanish is not that simple. In fact, authors disagree with respect to what is the basic word order in Spanish. Recent studies on Spanish word order claim that the basic word order of Spanish is VSO (Ordoñez and Treviño 1995, Zubizarreta 1995):

(4) *Spanish:*
Comió Juan los guisantes.
Ate Juan the peas.

In this paper, I will not take a position concerning whether one of the two descriptions is correct. Instead, I will assume that the two descriptions correspond to dialectal variation and that the two word orders are allowed as basic by different speakers. Henceforth, I will name the SVO variant Spanish A, and the VSO variant Spanish B.

Greek is another language with basic VSO order, as described in Alexiadou and Anagnostopoulou (1995):

(5) *Greek:*
Pandreftike o Petros tin Ilekttra
Married the-Peter-NOM the-Ilektra-ACC
'Peter married Ilektra'

Greek also permits SVO orders, but those are arguably analyzed in terms of left-dislocation of the subjects (see Alexiadou and Anagnostopoulou 1995 for details):

(6) *Greek:*
O Petros pandreftike tin Ilektra.
The Peter married the Ilektra

Arabic is quite similar to Greek in that its basic word order is VSO (SVO being derived via left-dislocation of the subject, as argued in Fassi-Fehri 1989 and Ouhalla 1991):

(7) *Arabic (from Ouhalla 1991):*
Sa-ya-shtarii Zayd-un dar-an.
Fut-3ms-buy Zayd house
'Zayd will buy a house'

Other VSO languages include Berber (8), Chamorro (9), and Celtic (10 and 11) (all examples are taken from Ouhalla 1991):

(8) *Berber:*
Ad-y-segh Moha ijn teddart.
Fut-3ms-buy Moha one house
'Moha will buy a house.'

(9) *Chamorro:*
Ha-taitai si Maria i lepblu.
3sg-read Maria the book.
'Maria read the book.'

(10)*Welsh:*
Gwelodd y bechgyn y draig.
Saw the boys the dragon
'The boys saw the dragon'

(11)*Irish:*
Deireann siad o paidir
Say they a prayer
'They say a prayer'

Note that in this introduction I am collapsing different types of VSO languages. In the following sections I will note some differences between them, which have led some authors to analyze them as belonging to distinct typological groups (cf. Ouhalla 1991 and Alexiadou and Anagnostopoulou 1995).

Dutch exemplifies the SOV group (though this word order only emerges in embedded context, its status as unmarked is the same with respect to information structure, see Dik 1978):

(12)*Dutch:*
dat Jan de krant leest.
That Jan the newspaper reads
'...that Jan reads the newspaper.'

VOS languages are illustrated by Malagasy (see Keenan 1976). In this language, the subject precedes both the verb and the object:

(13)*Malagasy:*
Nahita an-dRabe Rakoto
Saw ACC-Rabe Rakoto
'Rakoto saw Rabe'

In this paper I will not consider languages in which the O precedes both the subject and the verb in the unmarked case. Such languages are rare (cf. Derbyshire 1977 for OVS in Carib), and there is debate regarding their status as neutral in the attested cases (see Givón 1984 for discussion). I will nevertheless make an attempt to derive their rarity.

This variation at the base raises several questions, of which I would like to consider the following subset:

a) What does it (formally) mean to be the unmarked word order?
b) Why is there variation at the base?
c) Is the unmarked word order truly basic or is it in itself derived from a more basic structure?

The first question relates to the old question of what it means to be unmarked. There are several intuitive definitions of markedness (see Battistella 1996 for a review), but is there a clear formalism for expressing unmarkedness? In this paper, I will argue, along the lines of McCarthy and Prince (1994), that Optimality Theory (Prince and Smolensky 1993) provides such a formalism.

The second question is of a quite general nature: if all these word orders are felt as basic in the respective languages, why should there be different basic structures? That is, why is there not a uniform basis? The stand to be taken in this paper is that there may be a uniform basis from which these basic word orders are derived. The fact that different orders emerge as unmarked will follow from the constraint profile for each language, and consequently, the constraint with respect to which each order is unmarked.

Following the reasoning that underlies the answer to the second question, the position taken in this paper is that the basic word order does not necessarily correspond to an underived representation. That is, once we assume that the surface base representation may already be the result of compliance to a set of constraints, we may also assume (and in fact, must, in OT) that the identified word orders are themselves derived from a single uniform input. This is in accordance with the tradition and with more recent claims concerning word order, like Kayne (1994). Throughout the paper, I will not take a strong position with respect to whether there should be only one base word order (like Kayne does), but I would like nevertheless to stress the point that if this type of approach proves to be on the right track, one should not take the variation at the base as an argument for several basic word orders.

3. Theoretical Background: Optimality Theory and Emergence of the Unmarked.

The framework to be used in this paper is, as mentioned, Optimality Theory (Prince and Smolensky 1993). According to this theory, a particular grammar is a ranking of a set of universal violable constraints. These constraints evaluate the relative well-formedness of a set of candidates generated from a single input. Violation of one constraint does not imply ungrammaticality, provided that the other candidates violate higher ranked constraints. (T1) exemplifies a potential OT-diagram:

(T1)

	A	B	C
☞Candidate 1	*		*
Candidate 2	*	*!	

Tableau 1 shows that, in spite of the violation of constraints A and C, Candidate 1 is the grammatical one, since the other candidate fatally violates Constraint B, which is higher ranked than C. This tableau also shows that it does not matter how many violations there are. Ranking is the crucial factor. This diagram could be bigger, and include constraints D,

E and F, all violated by Candidate 1 and satisfied by Candidate 2, which would nevertheless be dispreferred. In a potential language where C dominates B the winning output candidate would be Candidate 2.

This type of theory is different from one where constraints are absolute and either active or inactive for each language. In fact, it predicts that the effects of a constraint which are normally invisible because all candidates which satisfy it fatally violate a higher ranked constraint may be visible in a context where the violation of the higher ranked constraints is controlled for. That is, the theory predicts that low-ranked constraints may be operative in some cases. The possibility of seeing the effects of a constraint that normally is violated is called by McCarthy and Prince (1994) **The Emergence of the Unmarked**.

This notion will be crucial in this article, since I will argue that unmarked or base word orders are just a subcase of Emergence of the Unmarked. The next section shows this for Portuguese.

4. Deriving the unmarked SVO order in Portuguese.

In this section I will show that the base word order of a language may indeed be explained in terms of Emergence of the Unmarked.

In Costa (1996), I have proposed that word order variation in Portuguese and its relation to discourse may be captured by assuming the constraints in (14) and the ranking given in (15):

(14) ***Constraints:***

a) ALIGNFOCUS: A focused constituent is the rightmost in a sentence.
b) SUBJ-CASE: Nominative Case is assigned in Spec,IP
c) OBJ-CASE: Accusative Case is assigned in Spec,AgrOP.
d) STAY: Don't move.
e) TOP-FIRST: Topics are sentence-initial (inspired in Li 1976). Failed by topics that are not sentence initial, by sentence initial non-topics (e.g. subjects that are not topics, and by topicalized non-topics.

These constraints are independently motivated and have been used in different domains, though not necessarily as violable constraints (cf. Grimshaw and Samek-Lodovici 1995,1996 for ALIGNFOCUS, Chomsky 1995 and Vikner 1997 for CASE, Grimshaw 1997 for STAY).

These constraints are in conflict, since STAY plays against movement, and CASE and ALIGNFOCUS may force a constituent to move. Likewise, ALIGNFOCUS and CASE are in conflict, since ALIGNFOCUS requires that a focused subject be rightmost in a sentence, while CASE requires that it move leftwards to Spec,IP. Similarly, TOP-FIRST conflicts with OBJ-CASE by forcing an object to surface in the topic position, instead of the

case-licensing position. In OT, these conflicts are resolved language-internally, in accordance to the ranking of constraints that constitutes the language's grammar.

For Portuguese, I proposed in Costa (1996) that the following ranking may characterize the behavior of Portuguese word orders:[1]

(15)***Ranking for Portuguese:***
{ALIGNFOCUS, TOP-FIRST} >>SUBJ-CASE>> STAY>>OBJ-CASE

This ranking determines the following: Subjects only move to Spec,IP if they are not focused; objects never move to Spec,AgrOP: they stay in their base position if they are focused and move, adjoining to VP if they are not focused. That this is the behavior of arguments in Portuguese has been argued for in Costa (1997,1996a). For other approaches to similar word order alternations, see Ambar (1992), Duarte (1987), Barbosa (1995,1996), Martins (1994), among others. For details concerning the implementation of the Optimality theoretical approach, see Costa (1996). The top-ranking of ALIGNFOCUS and TOP-FIRST reflects the discourse-configurational character of the language.

One case considered in Costa (1996) is what happens in cases of sentence focus. That is, what is the word order observed in contexts of answer to *what happened?* . As (16) illustrates, the emerging word order is SVO:

(16)O que é que aconteceu?
What happened
a. A Maria partiu um prato.
b. Maria broke a plate
c. #Partiu a Maria um prato.
d. #Partiu um prato a Maria.
e. #*A Maria um prato partiu.
f. #Um prato a Maria partiu.
g. #Um prato partiu a Maria
h. 'Maria broke a plate'

The argument for the reason behind the emergence of SVO as unmarked goes as follows: from a functional point of view, all these orders are equivalent, provided that the rightmost constituent bears the main stress (cf. Nespor and Vogel 1986, Frota 1994, among others). As long as focus projection is permitted (independently of its formulation in terms of syntactic constituency or in terms of linear order, see Costa 1996b for discussion), the whole sentence may be interpreted as focus independently of the word order displayed. For the hearer to interpret the whole sentence as focus, it is enough not to hear any high stress before the end of the sentence. When s/he hears an SVO sentence with main prominence on the object, s/he may interpret it as focus on the object, on the VP or on the whole sentence, depending which constituent is taken as the domain of focus. The same holds for a VSO

sentence with main prominence on the object: when the hearer gets it, the domain of focus may be the object alone, the subject and the object or the whole sentence.

Similarly, since there is no topic in these sentences, TOP-FIRST is violated by all word orders, since no sentence initial element will be the topic. Moreover, if some element is left-dislocated, this constraint is violated once more: one mark is due to the fact that the sentence initial element is not the a topic, and the other one because there was left-dislocation of a non-topic element. The crucial point is that in sentence focus contexts, all possible word orders violate TOP-FIRST.

The question is then why aren't all these word orders optional actualizations of sentence focus?

I would like to argue that this is where the Emergence of the Unmarked plays a crucial role.

By virtue of ALIGNFOCUS and TOP-FIRST being top-ranked, their effects very often make the requirements of the other constraints invisible. For instance, every time a subject is in focus, it is right aligned and consequently CASE has to be violated. If ALIGNFOCUS is not violated by any candidate, as it is the case in the sentence-focus context, the effects of Subj-CASE become visible. Since this is the next constraint, it is now the one by which decisions may be made, the winning candidate will be one which is not marked for this constraint. This is exemplified in (T2):

(T2)

	ALIGNFOCUS	TOP-FIRST	SUBJ-CASE	STAY	OBJ-CASE
☞a.[$_{IP}$ S V [$_{VP}$ t t O]]		*		**	*
b.[$_{IP}$ S V [$_{AgrOP}$ O [$_{VP}$ t t t]]]		*		***!	
c.[$_{IP}$ V [$_{VP}$ S t O]]		*	*!	*	*
d.[$_{IP}$ V [$_{AgrOP}$ O [$_{VP}$ S t t]]]		*	*!	**	
e.[$_{CP}$ O [$_{IP}$ S V [$_{VP}$ t t t]]]		**!		***	*
f.[$_{CP}$ O[$_{IP}$ V [$_{VP}$ S t t]]]		**!	*	**	*

The candidates considered in (T2) include:

- Candidate a.: SVO sentence with subject in Spec,IP and object in situ;
- Candidate b: SVO sentence with subject in Spec,IP and object in Spec,AgrOP;
- Candidate c: VSO sentence with subject in Spec,VP and object in situ;
- Candidate d: VOS sentence with subject in Spec,VP and object in Spec,AgrOP;
- Candidate e: OVS sentence with subject in Spec,IP and object in Spec,CP
- Candidate f: OVS sentence with subject in Spec,VP and object in Spec,CP.

The reader may have noticed that I have deliberately left out of consideration some word orders and representations. SOV order was left out, since I believe that there is an alternation at the base between head-initial and head-final languages. I also suppose that this alternation is to be analyzed at a deeper level than the one proposed here. Since SOV relates with directionality of theta-role assignment (Hoekstra 1984, Travis 1984, among others), and thematic information in OT is partially expressed in the input (cf. Grimshaw 1997), one has to work out a way of establishing the link between the information in the input and its correlation with the directionality of assignment. Such goal is beyond the scope of this paper.

Note nevertheless that it is conceivable to incorporate a more recent analysis of SOV word order into the present approach. Zwart (1993), following Kayne (1994), proposes that SOV languages are derived from SVO with overt movement of the object to Spec,AgrOP. This would mean compliance to Subj-CASE, Obj-CASE, two violations of STAY, and some violation of the constraint that forces V to move to I.

For reasons of coherence with previous work (Costa 1996), I will not adopt the latter hypothesis, stressing nevertheless that this is not crucial for the present paper.

The set of representations left out are all those in which the verb has not moved to I. The option of not including was taken for brevity and clarity purposes: it avoids adding constraints on V-to-I to the tableaux, making their reading much easier. For an analysis of V-to-I within OT, see Vikner (1997). Again, including those candidates would not change the argument to be developed in this paper.

Let us then proceed with a detailed analysis of each candidate, starting with the ungrammatical ones:

Candidate b. vacuously satisfies ALIGNFOCUS (like all the others). Since the Subject is in Spec,IP, Subj-CASE is satisfied. Satisfying Obj-CASE implies that the Object also moves out of VP, incurring in one mark for STAY.[2] This makes this candidate worse than candidate a. for STAY, the decision being made by this constraint.

Candidate c is ruled out, because the subject does not move to Spec,IP, violating Subj-CASE. This is a case of Emergence of the Unmarked: in general, violation of Subj-CASE is allowed in Portuguese, if the subject is focus. In that case, the effects of Subj-CASE are not visible. Here, since ALIGNFOCUS is vacuously satisfied by all candidates, the effects of Subj-CASE become apparent, and representations that do not cause this constraint to be marked are preferred.

Candidate d. is ungrammatical for the same reason. It fatally violates Subj-CASE.

Candidate e. and f. are slightly different, since they violate TOP-FIRST twice being immediately excluded. The reason for the second mark on TOP-FIRST is that not only the first element is not a topic (like in all the other candidates, but also there is an unnecessary topicalization of the object. This is maybe also penalized by a constraint not represented in the tableau playing against fronting of elements that are not operators (for constraints of this type, see Grimshaw 1997, Samek-Lodovici 1996, and Grimshaw and Samek-Lodovici 1995,1996).[3]

In the next section, I will show that minimal re-rankings between constraints of this set derive the unmarked word orders of the other languages discussed in the introduction.

5. Minimal re-rankings, base word orders and variation.

In this section, I will try to show that the variation at the base discussed above may be the result of the visibility of the effects of different dominated constraints, depending on the languages.

If the approach advocated here proves to be true, it has two important theoretical consequences, since it reinforces two of the premises of Optimality Theory:

First, it confirms the idea that constraints are violable. In fact, Emergence of the Unmarked is not expected in any approach defending absolute constraints, since a switched off principle/constraint/parameter should remain switched off independently of specific constructions. Effects of normally inactive constraints are only expected in a framework permitting violability. Moreover, in a model not contemplating constraint interaction as a legitimate possibility, it is difficult to explain why some principles are sometimes active and sometimes inactive.

Second, since one of the main goals of Optimality Theory is to provide a formalism to capture linguistic variation, the analysis here confirms this objective, by showing that re-ranking a constant set of constraints enables us to capture crosslinguistic differences.

5.1. Portuguese/Spanish A vs. Spanish B/Greek: SVO vs. VSO.

The first difference we will examine is the one between SVO and VSO of the type described for Spanish by Ordoñez and Treviño (1995) and for Greek by Alexiadou and Anagnostopoulou (1995). Since there is here variation among speakers of Spanish with respect to the difference between SVO and VSO, we do not want this difference to result from re-ranking of several constraints, since that might have consequences for the overall grammar of the language, which we would not be able to evaluate, and would predict differences between the two sets of speakers that are most likely untrue. Hence, in the optimal case, we expect the difference between the rankings for Spanish A and Spanish B to be minimal. If we manage to achieve that, we may also assume that the same differences are responsible for the difference between Portuguese and Greek.

Before presenting the crucial rankings, one has to examine what is the factor that makes these languages differ: according to the descriptions made in the literature, it seems to me that the behavior of subjects is the core of the difference.

According to most descriptions of Portuguese (Raposo (1987,1996), Mateus et alii (1989), Duarte (1987), Ambar (1992), Martins (1994), Costa (1997)),[4] preverbal subjects in Portuguese are in Spec,IP (or Spec,AgrSP). The same is argued for Spanish by Hernanz and Brucart (1987), who also describe Spanish as an SVO language. The argument these authors present for these descriptions is the behavior of preverbal subjects as A-moved

elements (not intervening for A-bar movement, behaving as A-binders, among other properties).

Differently, Alexiadou and Anagnostopoulou (1995) for Greek, and Ordoñez and Treviño (1995) for Spanish have proposed that these languages are VSO, in the sense that subjects remain in Spec,VP and preverbal subjects are just instances of subject left-dislocation. They argue in favor of this analysis on the basis of facts like: preverbal subjects either block A-bar movement (Spanish), or interacts with wh-phrases in the same way topics do (17); the interpretation of quantifiers is different depending on their position: only preverbal QPs have a strong reading (18); possibility of construing pronouns as bound variables, which is only possible with postverbal ones, showing that only the latter are in an A-position (19).

(17)*Greek* (from Alexiadou and Anagnostopoulou 1995):

a. Pjon (*o Petros) ide (o Petros)?
   Whom (the Peter-NOM) saw (the-Peter-NOM)

a'. *Pjos ton Petro ton ide?
   Who the Peter-ACC Cl-ACC saw
   'Who saw Peter?'

b. Pote (o laos) apofasise (o laos) na andidrasi?
   When (the people-NOM) decided (the people-NOM) SUBJ react
   'When did the people decide to react?'

b'. Pote tin tenia tin provalan ja proti for a
   When the movie-ACC Cl-ACC showed-3Pl for first time
   'When did they show the movie for the first time?'

(18)*Greek* (from Alexiadou and Anagnostopoulou 1995):

a. Enas heretise ti Maria. **Strong (partitive/specific) reading**
   one greeted the-Maria-ACC

b. 'A certain person/one of the people greeted Mary'
   Heretise enas ti Maria. **Weak (indefinite) reading**
   'Someone greete5d Mary'

(19)*Catalan* (from Solà 1992 and Alexiadou and Anagnostopoulou 1995):

a. *Tots els estudiants$_i$ es pensen que ells$_i$ aprovaran.
   All the students think that they will-pass
   'All the students think that they will pass'

b. Tots els jugadors$_i$ estan convencus que guanyaran ells$_i$.
   All the players are convinces that will-wing they
   'All the players are convinced that they are the ones who will win.'

Notice that, as mentioned before, Portuguese subjects are able to co-occur with fronted constituents, and wh-elements (20), and asymmetries like the one in (19) do not exist (21), both pre- and postverbal subjects may be construed as bound variables:

(20)*Portuguese:*

a. Que livros é que o João leu?
   Which books is that João read
b. Esses livros, o João leu.
   Those books, John read.

(21)*Portuguese:*

a. Todos os estudantes pensam que eles passarão.
   All the students think that they will pass
b. Todos os estudantes pensam que passarão eles.

On the basis of these asymmetries between the two sets of languages, I conclude that the distinction is finer than what is proposed in Alexiadou and Anagnostopoulou (1995), who propose that Null Subject Languages are all supposed to instantiate left-dislocation whenever a preverbal subject is found.

On the light of these data, it seems to me that Null Subject Languages do not behave uniformly: some have basic VSO word order and subjects do not move to Spec,IP (Greek and Spanish B), others have basic SVO word order and subjects may move to Spec,IP.

The conclusion is then that the crucial difference between these two sets of languages lies on whether preverbal subjects are A- or A-bar moved. Such difference has implications for which of the constraints presented above is satisfied.

Given the definition of the constraints given above, whenever a subject moves to Spec,IP, it satisfies Subj-CASE, at the expenses of violating STAY, since there is one movement operation. In a language in which Subj-CASE dominates STAY, it is more important to move the subject than satisfying economy. This is represented in (T3) below:

(T3)

	Subj-CASE	**STAY**
X a. $[_{IP}$ S V $[_{VP}$ t O]]		**
b. $[_{IP}$ V $[_{VP}$ S t O]]	*!	*

As for left-dislocation I will follow the various authors who claim that left-dislocation does not involve movement but base-generation of the left-dislocated element in adjunction to CP (see Duarte 1987, Cinque 1990, Raposo 1996, 1997). In this article, I will follow Raposo's (1996,1997) claim that all instances of topicalization and left-dislocation must be

treated as cases of base-generation.[5] Raposo argues that such an approach permits a unification of constructions of topicalization and null object constructions, as exemplified in (22):

(22)*Portuguese* (from Raposo 1996):
a. Muito whisky, comprei para o capitão.
A lot of whisky I bought for the captain
b. Comprei ec para o capitão.

Raposo claims that, once we acknowledge the necessity of having a null operator licensing the null object construction, as in his (1986) analysis following Huang (1984), we may as well unify the two constructions, by assuming that the two of them are licensed by a null operator, the difference being that the topic is overtly realized in (22a) but not in (22b).

Accepting this idea that all topicalizations involve base-generation of the topic, we are able to generalize over preverbal subjects in Spanish B and Greek, and assume that preverbal subjects in these languages are base-generated in a topic position (Spec,CP or adjunction to IP, depending on the analysis). That this is indeed the case is shown by Barbosa (1996), who shows that some preverbal subjects must associate with subject clitics overtly realized in Spec,IP in some Romance languages:

(23)*Trentino* (from Barbosa 1996):
a. El Mario *(el) parla
the Mario he speaks
b. Ti *(te) parli
You you speak

If preverbal subjects in these languages are base-generated in topic position, a crucial difference between them and preverbal subjects in Spec,IP is whether or not there is movement involved. Thus, the option of base-generating the subject in this position is more economical, which in our terms means that it does not violate STAY.

Whether a language chooses between moving the subjects to Spec,IP or base-generating them in topic position is then a matter of whether it is more important for the language to satisfy STAY (24) or Subj-CASE (25). The two possible situations are given in (T4) and (T5), where candidate a. is the one with subject left dislocation and candidate b. is the one with movement of the subject to Spec,IP:[6]

(24)STAY >> Subj-CASE

(25)Subj-CASE >> STAY

(T4) *Greek and Spanish B – Context: subject is the topic of the sentence*

	STAY	Subj-CASE
☞a. [$_{CP}$ S [$_{IP}$ pro V [$_{VP}$ t O]]]	*	*
b. [$_{IP}$ S V [$_{VP}$ t t O]]	**!	

(T5) *Portuguese and Spanish A – Context: subject is the topic of the sentence*

	Subj-CASE	STAY
a. [$_{CP}$ S [$_{IP}$ pro V [$_{VP}$ t O]]]	*!	*
☞b. [$_{IP}$ S V [$_{VP}$ t t O]]		**

(T4) reflects the situation in Spanish B and Greek, where it is more important to have an economical representation (without movement) than satisfying Case. In these languages, both candidates violate STAY once because of the movement of the verb. However, the candidate with movement of the Subject to Spec,IP involves one more violation of STAY, the second one being fatal.

(T5) represents what happens in Portuguese and Spanish A: satisfying Subj-CASE is more important than having an economical representation, hence the winner is the representation with movement of the subject to Spec,IP.

One may wonder why there is not a representation where the subject is base-generated in Spec,IP, in compliance with the two constraints. Here, I follow Grimshaw's (1997) claim that the only candidates generated by GEN, are those that comply with general principles of X-bar structure and the thematic criterion. According to Koopman and Sportiche (1991), the thematic role of the subject is assigned under sisterhood to V-bar, hence base generation of the subject in Spec,IP would yield a violation of the theta-criterion, precluding generation of such candidate.[7] Note that base-generation of *pro* in Spec,IP is not problematic, since this is a clitic-like pronominal element which is very often assumed to be the instantiation of a functional projection (see Zwart 1993, Kayne 1991), hence described as an agreement marker and not as the argument of the verb, able to receive Case or thematic role. Naturally, this type of analysis raises questions that fall beyond the scope of this article, related to how the thematic interpretation of the left-dislocated subject takes place after all. In principle, this issue may be solved within the theory, since the thematic information may be not expressed by a local relation between theta-role assigner and assignee in the output, but it will always be represented in the input. If the input contains all the semantic information (see Grimshaw 1997, Légendre et alii 1995, Costa forthcoming for discussion), it does not follow in any way that a subject which does not get its theta-role in a local relation with the assigner will not be interpreted as the thematic subject of the verb.

So far, we have just represented the choice between SVO with movement to Spec,IP and SVO with left-dislocation of the subject. We still have to explain why for these two language groups the unmarked word order is SVO and VSO, respectively.

The crucial part of the explanation is again the ranking between Subj-CASE and STAY, though TOP-FIRST also plays a role. We are now going to consider three candidates: SVO with left-dislocation, SVO with subject in Spec,IP, and VSO. Since the context is sentence focus, all orders will violate TOP-FIRST once, for the first element of the sentence is not a topic. In addition, the candidate involving left-dislocation of the subject will violate it once more, since there is topicalization of a non-topic element. This is similar to the way we excluded OSV and OVS as potential base word-orders. The prediction is then that there should be no unmarked word order involving left-dislocation of any type. Having controlled the effects of ALIGNFOCUS and TOP-FIRST, we may now see whether the choice is made by Subj-CASE or by STAY. As expected, in Portuguese and Spanish A, it will be more important to satisfy CASE than STAY, hence subjects will move to Spec,IP. This is represented in (T6). In Greek and Spanish B, it is more important to satisfy STAY than Subj-CASE, hence subjects will not move to Spec,IP, yielding a more economical representation. This is schematized in (T7):

(T6) *Portuguese and Spanish A – Context: sentence-focus*

	ALIGNFOCUS	TOP-FIRST	SUBJ-CASE	STAY	OBJ-CASE
a. [$_{CP}$ S [$_{IP}$ V [$_{VP}$ t O]]]		**!	*	*	*
☞b. [$_{IP}$ S V [$_{VP}$ t t O]]		*		**	*
c. [$_{IP}$ V [$_{VP}$ S t O]]		*	*!	*	*

(T7) *Greek and Spanish B – Context: sentence-focus*

	ALIGNFOCUS	TOP-FIRST	STAY	SUBJ-CASE	OBJ-CASE
a. [$_{CP}$ S [$_{IP}$ V [$_{VP}$ t O]]]		**!	*	*	*
b. [$_{IP}$ S V [$_{VP}$ t t O]]		*	**!		*
☞c. [$_{IP}$ V [$_{VP}$ S t O]]		*	*	*	*

It is important to note that, in spite of the relevance of TOP-FIRST for the exclusion of the candidate with left-dislocation of the subject, it is the ranking between STAY and SUBJ-CASE that is crucial. The decision between candidates b. and c. is made at the point one of these two constraints (depending on the ranking) evaluates the candidates. In (T6), SVO is selected, because only in this order, the subject has moved to Spec,IP, satisfying Subj-CASE. In (T7), because of the reverse ranking, STAY plays against SVO, since there are two traces in the structure as opposed to one in the VSO order.

Note that we have achieved one of our purposes in capturing the difference between SVO and VSO: given the difference between Spanish A and Spanish B, which is dialectal,[8] we would not like to have a huge difference in constraint profiles. Indeed, the difference between the two languages is minimal: only SUBJ-CASE and STAY need to be re-ranked with respect to one another. (26) and (27) are the rankings for Portuguese and Spanish A, and Greek and Spanish B, respectively:[9]

(26)*Portuguese and Spanish A:*
{ALIGNFOCUS, TOP-FIRST} >> **Subj-CASE >> STAY** >> Obj-CASE

(27)*Greek and Spanish B:*
{ALIGNFOCUS, TOP-FIRST} >> **STAY >> Subj-CASE** >> Obj-CASE

In this way, we manage to keep under control the consequence of re-ranking these constraints for other aspects of the grammar of the languages under consideration, since Subj-CASE has in its scope a quite restricted set of elements. If we were to radically modify the constraint profile, we would be very likely to cause differences in between the two dialects of Spanish that are not real.

5.2. Postverbal subjects: VSO vs VOS.

In the preceding section, we have derived the difference between SVO and VSO as unmarked word orders for Portuguese and Spanish A and Greek and Spanish B, respectively. In this section, I would like to develop the analysis presented there in order to accommodate a subtle difference between Portuguese and another Romance language: Italian.

Italian is like Portuguese, in that its unmarked word order is SVO, as (28) illustrates:

(28)*Italian* (from Pinto 1997):
Che cosa è sucesso?
What happened?
a. Beatrice ha scritto lettere d'amore.
Beatrice has written love letters
b. #Ha scritto lettere d'amore Beatrice.

In her description of neutral word orders, Pinto notes, following Calabrese (1991) among others, that some inverted structures are felicitous as answers to *what happened.* That is the case with verbs that select some kind of locative argument. Though Pinto's work focuses mainly on such constructions, I will overlook them here, for they are not possible with transitive verbs, thus irrelevant for the present discussion, since it becomes impossible to test the ordering between subject and object. I refer the reader to Pinto (1997) for a description of the contexts of VS order in Italian. The possibility of having VS

orders with such verbs may be derived from restrictions on the discourse situation imposed by the lexical meaning of the verbs or by the locative argument itself.

The crucial aspect where Italian differs from Portuguese is that though VOS orders are allowed when only the subject is in focus (29), VSO orders which are legitimate in Portuguese are ungrammatical in Italian, independently of the context (30) (also noted by Rizzi 1982, Burzio 1986, among others):

(29)*Italian* (from Pinto 1997):
Chi ha scritto lettere d'amore?
a. #Beatrice ha scritto lettere d'amore.
Beatrice has written love letters
b. Ha scritto lettere d'amore Beatrice.

(30)Nessuno ha scritto niente.
No-one has written anything
a. *Ha scritto Beatrice lettere d'amore.
b. Ha scritto lettere d'amore Beatrice.

Remember that the context in (30) is the one that legitimates VSO orders in Portuguese:

(31)Ninguém escreveu nada.
a. Escreveu a Beatriz cartas de amor.
b. ?#Escreveu cartas de amor a Beatriz.

Here I will follow Pinto's analysis of VOS order. Pinto claims that in spite of the lack of evidence from adverbial placement (cf. Belletti 1990), VOS orders may be analyzed as an instance of subject in Spec,VP with movement of the object to Spec,AgrOP, as in the representation in (32):

(32)$[_{IP}$ V $[_{AgrOP}$ O $[_{VP}$ S t t]]]

Such an analysis is also defended in Cinque (1997), who argues for this on theoretical grounds by rejecting right adjunction as a possibility, and by the observation that postverbal subjects in Italian transitive constructions must be absolutely final. Actually, the observation by Pinto that inverted subjects may precede VP-adjuncts (cf. 33) confirms this hypothesis, since we would otherwise expect to find right dislocated subjects following all base-adjuncts:

(33)*Italian* (from Pinto 1997):

a. E' arrivato Dante da Firenze.
   Is arrived Dante from Florence
   'Dante has arrived from Florence.'
b. Ha telefonato Beatrice da Milano.
   Has called Beatrice from Milano
   'Beatrice has called from Milano.'

The only possibility to rescue the VP-adjoined position for subjects in sentences like (33) would be to assume a multiple right-adjunction structure for the VP, which would nevertheless be problematic, since as Samek-Lodovici (1996) argues, this position (right adjunction to VP) appears to be reserved for contrastive foci, upon which a restriction of uniqueness applies (cf. Szabolcsi 1981).

I will therefore follow Pinto (1997) and Cinque (1997) for the structural representation of VOS order in Italian.[10]

The question now is whether we can represent the situation in Italian by playing with the set of constraints used before. Before starting trying possible rankings, it is important to understand where the crucial difference between Italian and Portuguese is. What seems to be at stake is that Italian has obligatory object-movement to Spec,AgrOP, while Portuguese does not have it at all.[11] As the examples above illustrated, not even context may force a VOS word order in Italian. In our system, this means that the constraint driving the movement of the object is more important than those related to discourse. That is, the ranking between OBJ-CASE and ALIGNFOCUS has to be the one in (34):

(34)OBJ-CASE >> ALIGNFOCUS

Since subjects do not behave this way (their position is dependent on the discourse context), the ranking between ALIGNFOCUS and SUBJ-CASE has to be kept the same as for Portuguese:

(35)ALIGNFOCUS >> SUBJ-CASE

Since preverbal subjects are in Spec,IP, according to most analyses of Italian (Rizzi 1982, Burzio 1986, Belletti 1990, Samek-Lodovici 1996, Pinto 1997, among many others), it may not be the case that STAY dominates Subj-CASE (which also derives the impossibility of having scrambling as adjunction to VP in this language).

(36)SUBJ-CASE >> STAY

Since topicalization is possible in Italian (cf. Rizzi 1995 among others), TOP-FIRST has to be more important than OBJ-CASE.

(37)TOP-FIRST >> OBJ-CASE

We arrive thus at the following partial ranking for Italian:

(38)TOP-FIRST >> OBJ-CASE >> ALIGNFOCUS >> SUBJ-CASE >> STAY

Though it seems quite different from the ranking for Portuguese presented in (39), notice that once again we only had to re-rank one constraint (OBJ-CASE). This is highly desirable, since we are still talking about variation within one family of languages.

(39){ALIGNFOCUS, TOP-FIRST} >> SUBJ-CASE >> STAY >> OBJ-CASE

Let us now see how the selection of candidates is done. For this case, we have to consider a context where both the subject and the object are focused. (T8) represents the evaluation tableau for Portuguese, where VSO wins. (T9) is the evaluation tableau for Italian, where VOS wins. The VSO candidate and the VOS candidates with movement of the object to Spec,AgrOP and with scrambling of the object in adjunction to VP are included in the set of candidates under comparison.

(T8)

	ALIGNFOCUS	TOP-FIRST	SUBJ-CASE	STAY	OBJ-CASE
☞a. $[_{IP}$ V $[_{VP}$ S t O]]			*	*	*
b.$[_{IP}$ V $[_{AgrOP}$ O t $[_{VP}$ S t t]]]	*!		*	***	
c. $[_{IP}$ V $[_{VP}$ O $[_{VP}$ S t t]]]	*!		*	**	*

(T9)

	TOP-FIRST	OBJ-CASE	ALIGNFOCUS	SUBJ-CASE	STAY
a. $[_{IP}$ V $[_{VP}$ S t O]]		*!		*	*
☞b.$[_{IP}$ V $[_{AgrOP}$ O t $[_{VP}$ S t t]]]			*	*	***
c. $[_{IP}$ V $[_{VP}$ O $[_{VP}$ S t t]]]		*!	*	*	**

As can be seen in the tableaux above, the decision for which candidate is the optimal one is made quite early in the evaluation diagram, in the sense that it is done by one of the top-ranked constraints. In this sense, one might think that the ranking of SUBJ-CASE and STAY in Italian is irrelevant, since OBJ-CASE immediately filters out all but one candidate. However, once again the effects of Emergence of the Unmarked have to be considered here. Only the ranking proposed may accommodate the facts of postverbal

subjects and the emergence of SVO order in the unmarked case. This is shown in (T10), where the following candidates are considered:

(40)Candidate a: SVO with Subject in Spec, IP and Object in Spec,AgrOP;
Candidate b: SVO with Subject in Spec,IP and Object in situ;
Candidate c: SVO with Subject in Spec,IP and scrambled object;
Candidate d: SVO with Subject right-dislocated and Object in Spec,AgrOP;
Candidate e: SVO with Subject right-dislocated and Object in situ;
Candidate f: SVO with Subject right dislocated and scrambled object;
Candidate g: VSO with Subject and Object in situ;
Candidate h: VOS with Subject in Spec,VP and Object in Spec,AgrOP;
Candidate i: VOS with Subject in Spec,VP and Scrambled object.

(T10)

	TOP-FIRST	OBJ-CASE	ALIGNFOCUS	SUBJ-CASE	STAY
☞a. [$_{IP}$ S V]$_{AgrOP}$ O t [$_{VP}$ t t t]]]	*				****
b. [$_{IP}$ S V]$_{VP}$ t t O]]	*	*!			**
c. [$_{IP}$ S V [$_{VP}$ O [$_{VP}$ t t t]]]	*	*!			***
d. [$_{CP}$ S [$_{IP}$ V]$_{AgrOP}$ O t [$_{VP}$ t t]]]	**!			*	***
e. [$_{CP}$ S [$_{IP}$ V]$_{VP}$ t O]]]	**!	*		*	*
f. [$_{CP}$ S [$_{IP}$ V]$_{VP}$ O [$_{VP}$ t t]]]	**!	*		*	**
g. [$_{IP}$ V [$_{VP}$ S t O]]	*	*!		*	*
h. [$_{IP}$ V [$_{AgrOP}$ O t [$_{VP}$ S t t]]]	*			*!	***
i. [$_{IP}$ V [$_{VP}$ O [$_{VP}$ S t t]]]	*	*!		*	**

Let us recap what happens in (T10): This tableau shows that SUBJ-CASE and STAY, though dominated by three constraints, are active. Let us look at the place where candidates fatally violate one constraint: all the candidates with left-dislocation of the subject fatally violate TOP-FIRST, since they involve topicalization of a non-topic (similarly to what happens in Greek). OBJ-CASE filters out all candidates with the object in situ or scrambling, precluding the VSO order to emerge. A decision has to be made now between SVO and VOS (candidates a. and h. respectively). ALIGNFOCUS does not play any role, hence SUBJ-CASE will decide for the candidate that satisfies it: SVO. The VOS order fatally violates SUBJ-CASE. Note that here we may also observe the crucial ranking between SUBJ-CASE and STAY. If STAY would dominate ALIGNFOCUS, the VOS order would emerge, since it only involves three movement operations, against the four movements involved in the representation a.

Summing up, once again we managed to describe a language by virtue of a minimal re-ranking of the assumed constraints, and to show that the effects of dominated constraints are visible in sentence-focus context.

5.3. VOS in Italian vs. VOS in Malagasy.

The analysis proposed for Italian may be extended to accommodate the VOS word order that emerges as unmarked in Malagasy, as (41) illustrates:

(41)*Malagasy (from Keenan 1976):*
Nividy mofo ho'an'ny ankizy aho
Bought bread for the children I
'I bought bread for the children.'

Differently from Italian, this word order is rigid: complement of the verb may not follow the subject:

(42)a. *Nividy aho mofo ho'an'ny ankizy
bought I bread for the children
b. *Nividy ho an'hy ankizy aho mofo.
Bough for the children I bread

Likewise, SVO orders are not allowed (unless there is some kind of cleft construction or topicalization, which are morphologically marked (see Keenan 1976)):

(43)*Aho nividy mofo ho'an'ny ankizy.
I bought bread for the children

Evidence for analyzing this word order in the same terms as the analysis proposed for Italian, repeated in (44), comes from the distribution of question, exclamative and no-longer particles. These particles are often used to determine the position of the subject (see e.g. Diesing 1992 for German). In this language, they always precede the subject:

(44)$[_{IP}$ V $[_{AgrOP}$ O $[_{VP}$ S t t]]]

(45)a. Nanome vola an-dRabe ve ianao?
Gave money acc-Rabe Q you
'Did you give money to Rabe?'
b. Manasa lamba anie Rasoa!
Washes clothes EXCL Rasoa
'Is Rasoa ever washing clothes!'
c. Tsy manasa lamba intsony Rasoa.
Not washes clothes longer Rasoa
'Rasoa is no longer washing clothes.'

From the examples above, one may hypothesize that subjects in Malagasy are VP-internal.[12]

Expressing this in terms of ranking, we have to capture the fact that VOS is the unmarked word-order. We can get that by taking the same ranking for Italian and changing the relative ranking of STAY and Subj-CASE. For Malagasy, the ranking between these two constraints must be:

(46) STAY >> SUBJ-CASE

This predicts that it is more important for a representation to be economical than for an NP to be assigned nominative Case in Spec,IP.

This, however, cannot be all. The discourse-related constraint ALIGNFOCUS must be relatively low-ranked in this language, since different contexts do not permit different word orders, unlike in the other languages considered in this paper. Since subjects are always sentence-final, I propose that ALIGNFOCUS is ranked below SUBJ-CASE. I will not make any proposal concerning TOP-FIRST: since there are topicalizations in this language, TOP-FIRST must be relatively high-ranked. However, I do not know how it interacts with the other constraints. For the sake of exposition, I will leave it top-ranked, but the reader should be aware that this may be wrong. Additional constraints related to morphologically marked topicalization, and focus-related constraints imposing requirements on the position of contrastively focused elements will most likely be ranked above the constraints discussed here, and trigger morphologically marked topicalization and clefting.

Let us now look at the evaluation tableau (T11), and see how the ranking proposed derives the word order in Malagasy:

(T11)

	TOP-FIRST	OBJ-CASE	STAY	SUBJ-CASE	ALIGNFOCUS
a. [$_{IP}$ S V [$_{AgrOP}$ O t [$_{VP}$ t t t]]]	*		****!		
b. [$_{IP}$ S V [$_{VP}$ t t O]]	*	*!	***		
c. [$_{IP}$ S V [$_{VP}$ O [$_{VP}$ t t t]]]	*	*!	***		
d. [$_{CP}$ S [$_{IP}$ V[$_{AgrOP}$ O t [$_{VP}$ t t]]]	**!		***	*	
e. [$_{CP}$ S [$_{IP}$ V [$_{VP}$ t O]]]	**!	*	*	*	
f. [$_{CP}$ S [$_{IP}$ V [$_{VP}$ O [$_{VP}$ t t]]]	**!	*	**	*	
g. [$_{IP}$ V [$_{VP}$ S t O]]	*	*!	*	*	
☞h. [$_{IP}$ V [$_{AgrOP}$ O t [$_{VP}$ S t t]]]	*		***	*	
i. [$_{IP}$ V [$_{VP}$ O [$_{VP}$ S t t]]]	*	*!	**	*	

(T11) shows that the relevance of SUBJ-CASE, and ALIGNFOCUS is none for deriving the unmarked word order of this language. The decision is made by TOP-FIRST that rules out all candidates with left-dislocation of the subject, by OBJ-CASE that rules out all candidates without movement of the object to Spec,AgrOP, and by STAY which decides between candidate a. and candidate h., preferring the most economical representation: the one without movement of the subject.

This case is slightly different from the other languages discussed, since it is not the case that ALIGNFOCUS dominates everything else and that we need to look at sentence-focus contexts to see the effects of the other constraints. It is nevertheless worth noting that a rearrangement of the constraints derives the unmarked word order of this language. The role played by TOP-FIRST (if any, see discussion above) is however the same as the role played by ALIGNFOCUS in the other languages.

5.4. More on VSO: Celtic and Arabic.

The preceding discussion on Malagasy VOS as opposed to Italian VOS enables us to give an account of the opposition between VSO in Celtic versus VSO in Arabic, Chamorro and Berber. The difference between these languages is, according to Ouhalla (1991), that Berber, Chamorro and Arabic are VSO languages allowing SVO as a possible word order (namely when the subjects are topics), while Celtic never allows SVO: it is a rigid VSO language. This is similar to the alternation we found between SVO and VOS in Italian versus VOS in Malagasy, though the relation word order/function is different.

The following examples taken from Ouhalla (1991) illustrate these patterns for Berber, Arabic and Welsh:

(47)*Berber:*

a. ad-y-segh Moha ijn teddart.
   Fut (TNS)-3ms (AGR)-buy Moha one house
   'Moha will buy a house.'
b. Moha ad-y-segh ijn teddart.
   'Moha will buy a house.'

(47)*Arabic:*

a. Sa-ya-shtarii Zayd-un dar-an
   Fut (TNS)3ms (AGR)-buy Zayd-nom house-acc
   'Zayd will buy a house.'
b. Zayd-un sa-ya-shtarii daar-an.
   Zayd-nom fut (TNS)3ms(AGR)-buy house-acc
   'Zayd will buy a house;

(47)*Welsh:*

a. Gwelodd y bechgyn y draig.
   Saw the boys the dragon
   'The boys saw the dragon.'
b. *Y bechgyn gwelodd y draig.
   The boys saw the dragon.

Ouhalla's explanation for this difference, though interesting, may not be integrated here. Ouhalla notes that Welsh differs from the other languages not only in this, but also in two other aspects which are characteristic of SVO languages: Celtic languages have non-inflected infinitives and the order of inflectional morphemes is TNS-AGR. Ouhalla suggests that word order alternations (VSO enables SVO, but not vice versa), the order of morphemes and the availability of non-inflected infinitivals are three consequences of one parametric difference: whether in the clause structure AGR selects T or T selects AGR. Ouhalla shows that all the VSO properties follow from a structure where AGR selects T, and the SVO from a structure where AGR selects T. Under Ouhalla's analysis, in a language where T selects AGR, the subject ends up in Spec,AgrSP, not needing to move further up, though Spec,TP remains a legitimate position for topics. In the languages where AGR selects T, the subject has to move all the way up to AGR for Case purposes, yielding the SVO order. Celtic falls within the latter group of languages, with the language-particular rule saying that subjects are assigned nominative Case in Spec,VP.

The reason for not integrating this analysis is that it resorts to two mechanisms not available in OT: language-particular rules and parametrization of functional heads. In OT, constraints are universal, and all language variation has to follow from constraint ranking. Furthermore, we have been following the claims put forward in the literature for a dialect of Spanish that this language is VSO, though the language exhibits the order of morphemes TNS-AGR, a counter-example to Ouhalla's generalization.[13]

Let us first consider Arabic: this language will have an analysis very similar to the one proposed for Greek and Spanish B: VSO in the unmarked case is a consequence of the dominance of SUBJ-CASE by STAY. The dependence on discourse context will follow from the high-ranking of TOP-FIRST and ALIGNFOCUS. Hence, we will get exactly the same situation as proposed for Greek and Spanish B, represented in (T12):

(T12)

	ALIGNFOCUS	TOP-FIRST	STAY	SUBJ-CASE	OBJ-CASE
a. [$_{CP}$ S [$_{IP}$ V [$_{VP}$ t O]]]		**!	*	*	*
b. [$_{IP}$ S V [$_{VP}$ t t O]]		*	**!		*
☞c. [$_{IP}$ V [$_{VP}$ S t O]]		*	*	*	*

Departing from the assumption made above for Malagasy that rigid word order is a consequence of the low ranking of the discourse-related constraints, we may now explain why Celtic has a rigid VSO order. This order will emerge in context of sentence-focus, represented in (T13), as expected:

(T13)

	STAY	SUBJ-CASE	OBJ-CASE	ALIGNFOCUS	TOP-FIRST
a. [$_{CP}$ S [$_{IP}$ V [$_{VP}$ t O]]]	*	*			**!
b. [$_{IP}$ S V [$_{VP}$ t t O]]	**!		*		*
☞c. [$_{IP}$ V [$_{VP}$ S t O]]	*	*	*		*

Note that it is still TOP-FIRST that decides between candidate a. and candidate c. However, the low ranking of TOP-FIRST is important to make sure that the same word order will be optimal when the subject is not in focus. This is what happens in (T14): in this tableau it is necessary to resort to other structural constraints, so that we can see that it is really the effect of structural constraints that determine the VSO word order, and that the discourse-related constraints are very low-ranked. Hence, in (T14), I add *STRUC (from Grimshaw 1997), which plays against generation of phrase structure, and is fatally violated by the candidate with CP projected. If this constraint would not play a role, candidate a. would be the winner. The case of Celtic serves then as a proof to the issue raised in footnote 3 that topicalization should be controlled for by two types of constraints: structural and functional. Given the main claim made in this paper and in Grimshaw and Samek-Lodovici (1995) and in Costa (1996) that optionality arises when discourse constraints dominate syntactic constraints, while rigid word order arise when the reverse happens, it is not surprising that *STRUC dominates both ALIGNFOCUS and TOP-FIRST. If that would not be the case, candidate a. would win like in Greek and Spanish B.

(T14)

	STAY	SUBJ-CASE	OBJ-CASE	*STRUC	ALIGNFOCUS	TOP-FIRST
a. [$_{CP}$ S [$_{IP}$ V [$_{VP}$ t O]]]	*	*	*	***!		
b. [$_{IP}$ S V [$_{VP}$ t t O]]	**!		*	**		
☞c. [$_{IP}$ V [$_{VP}$ S t O]]	*	*	*	**	*	*

Note that adding *STRUC to the constraint profile is not a stipulation to derive the Celtic facts: this constraint is independently used in the work of Grimshaw, and the role it plays here confirms the idea that the rigid word order follows from the very low-ranking of ALIGNFOCUS and TOP-FIRST.[14]

By not adopting the whole of Ouhalla's analysis, we loose the relationship between weak and strong agreement and position of the subject. Ouhalla proposes that subject agreement correlates with the position of the subject at S-structure: Arabic and Celtic have weak Agreement because the verb is in Spec,VP, while Berber and Chamorro have rich agreement because the verb is in Spec,AgrSP. We do not loose an account of the agreement

patterns provided that we adopt the strong lexicalist hypothesis (Chomsky 1995), and as long as we make sure that agreement is accounted for somewhere else in the model: morphological alignment constraints may account for it (McCarthy and Prince 1994, Grimshaw 1997). Since Celtic and Arabic have a similar agreement pattern, it is not clear that there is a correlation between the typological groups Ouhalla proposes and richness of Agreement. In that case, we maybe loose the account of agreement in the syntax, but account for the position of the subjects in a uniform way and without resorting to a language-particular rule.

6. Conclusion.

In this paper, I have tried to derive the unmarked word orders of several languages in terms of emergence of the unmarked. It was shown the effect of constraints normally invisible by virtue of the domination by discourse-related constraints becomes visible in context of sentence-focus.

The following pairs ranking/language were proposed:

(50) a. *Portuguese and Spanish A:*
{TOP-FIRST,ALIGNFOCUS}>>Subj-CASE>>STAY>>Obj-CASE
b. *Spanish B, Greek, Arabic, Berber and Chamorro:*
{TOP-FIRST,ALIGNFOCUS}>>STAY>>Subj-CASE>>Obj-CASE
c. Italian:
TOP-FIRST>> Obj-CASE>> ALIGNFOCUS >> Subj-CASE >> STAY
d. Malagasy:
Obj-CASE >> STAY >> Subj-CASE >> TOP-FIRST >> ALIGNFOCUS
e. Celtic:
STAY>> Subj-CASE >> Obj-CASE >> *STRUC >> {TOP-FIRST,ALIGNFOCUS}

The rarity (or nonexistence) of OVS and OSV as unmarked word orders was explained in terms of the ban on topicalization in sentence-focus contexts, on a pair with the ban on left-dislocated subjects in such context.

The only word order not considered here was SOV for the reasons discussed in section 2, and because accounting for it would imply taking into account the constraints on V-to-I discussed in Grimshaw (1997) and Vikner (1997), which would make us go beyond the goal of this paper.

If the analysis presented here is on the right track, it is a step forward in showing the power of OT as a theory of language variation and constraint interaction, since minimal differences between languages may be express by means of minimal re-rankings between constraints.

NOTES:

1 Actually, in Costa (1996) I did not consider TOP-FIRST, since this constraint was not relevant there. As it will become evident, this constraint is only necessary when OS orders are considered.
2. The discussion here is quite simplified for reasons of exposition. Actually, candidate b. should have four marks for STAY, since satisfying Obj-CASE involves projecting AgrOP, which creates one more landing site for the verb. I am not including that mark for making the reading of the diagrams easier, and because I am leaving verb movement out of the discussion all together.
3. It is not unlikely that this constraint may have to be split up in two. There are several conceivable ways to express the same idea that topicalization of non-topics should not be permitted: not only it violates TOP-FIRST, as it is defined above, but it also constitutes a violation of *STRUC (a constraint penalizing structure, cf. Grimshaw 1997), since it involves one extra layer of structure. In this paper, I will mainly adhere to the idea that the reason for its exclusion is to be attributed to a constraint that is functional in nature. One advantage of such an approach is the derivation of the fact that OS word orders are very rare if not nonexistent as unmarked. This would follow from the fact that the top-ranked functional/discourse-related constraints would immediately filter out such sequences. Note that this is not a matter of choice between the structural and the functional approach, since both constraints are necessary, as it will become evident when we look at Celtic. Unless necessary (see discussion of Celtic below) I will not represent *STRUC in the tableaux.
4. See however Barbosa (1995) for arguments for a different view.
5. This approach differs from the tradition, where it is assumed that there must be a partition between topicalization (involving movement) and (Clitic) Left dislocation (not involving movement). This is for instance the distinction proposed in Cinque (1990). Duarte (1987) suggests that some topicalizations may involve A-bar movement of the topic. For discussion, see the works cited.
6. Note that VSO word order in this context is independently ruled out by ALIGNFOCUS, since this order forces the subject to be interpreted as focus, which is not the case: we are considering contexts in which the subject is the focus.
7. Note that there may be candidates with arguments unrealized and theta-roles undischarged. This option is permitted under Samek-Lodovici's (1996) formulation of the theta-criterion, which imposes requirements on locality relations between assigners and assignees **only when the theta-role is discharged**.
8. This statement must be qualified. I do not possess information enabling me to tell whether the difference between VSO and SVO in Spanish is dialectal or ideolectal. For the purposes of this paper, it is enough to say to note that there is variation among speakers, and that the differences between the rankings capturing it ought to be minimal.
9. The rankings proposed also predict that all these languages have scrambling of objects as A-bar movement (in adjunction to VP) (cf. Costa 1996), if they have scrambling. This is partially confirmed for Spanish by Ordoñez (to appear), who shows that object movement in Spanish is A-bar. Alexiadou and Anagnostopoulou (1996), in their comparative work on clitic doubling and scrambling show that Greek objects do not move to Spec,AgrOP (instead there is clitic doubling in this language). What it does not predict is that subjects and objects behave uniformly with respect to Case-licensing within the same language, as it is proposed in Chomsky (1993). The facts of Portuguese do not seem to confirm such proposal, hence I will not see this as a weakness of my analysis.
10. Note that I am only partially following Cinque, since he does not include Agr-phrases in the inventory of potential functional projections, following Chomsky (1995).
11. I have not presented the facts here, but there seems to be reasons to believe that Portuguese VOS orders are derived via adjunction to VP instead of movement to Spec,AgrOP, a fact explained in Costa (1996) by the domination of OBJ-CASE by STAY. The explanation goes as follows: by satisfying Obj-CASE, a representation is forced to project AgtOP which creates one more head position for the verb to land in on its way to I, yielding one extra violation of STAY, and rendering the representation less economical.
12. One problem with this analysis is how to explain the obligatory movement to AgrOP of non-nominal complements. At this stage, I do not have anything interesting to suggest. Another surprising factor concerning this analysis is the Definiteness restriction presented by Keenan: subjects must be definite. This is strange

vis-à-vis our conclusion, since it is more common for VP-internal subjects to be indefinite. Keenan presents data showing that whenever there is an indefinite subject, a special type of existential construction must be used. Probably, the availability of this special construction in the language blocks the use of indefinites in the canonical subject position.

13. Note that nothing prevents an extension of Ouhalla's analysis for Celtic to Spanish: Spanish might be an AGR-initial language with the language-particular rule assigning nominative Case in Spec,VP.

14. One possible analysis not involving *STRUC and still deriving the right results, would be to follow Sproat (1985), and assume that VSO in Celtic is an instance of I-to-C, which would make candidate a. in (T14) violate STAY twice. McCloskey (1992), Carnie (1995) and Bobaljik (1995) have, however, argued based on the distribution of complementizers that such an analysis is problematic. I will follow the V-to-I analyses, which permit a unification with the Greek and Spanish case.

REFERENCES:

ALEXIADOU, ARTEMIS AND ELENA ANAGNOSTOPOULOU. 1995. SVO and EPP in Null Subject Languages and Germanic. In *FAS. Papers in Linguistics*, Potsdam

ALEXIADOU, ARTEMIS AND ELENA ANAGNOSTOPOULOU. *Symmetries, Asymmetries and the Role of Agreement*. Paper presented at GLOW. XIX in Athens

AMBAR, MANUELA. 1992. *Para uma Sintaxe da Inversão Sujeito Verbo em Português*. Doctoral dissertation, University of Lisbon

BARBOSA, PILAR. 1995. *Null Subjects*. Doctoral Dissertation, MIT

BARBOSA, PILAR. 1996. A New Look at the Null Subject Parameter. In João Costa et alii (eds.) *Proceedings of ConSOLE 3*, Leiden University

BATTISTELLA. 1996. *The Logics of Markedness*. Oxford University Press

BELLETTI, ADRIANA. 1990. *Generalized Verb Movement*. Rosenberg and Sellier, Torino

BOBALJIK, JONATHAN. 1995. *Morphosyntax: the Syntax of Verbal Inflection*. Doctoral dissertation, MIT

BURZIO, LUIGI. 1986. *Italian Syntax*. Kluwer

CALABRESE, ANDREA. 1991. Some Remarks on Focus and Logical Structures in Italian. In *Harvard Working Papers in Linguistics*

CARNIE, ANDREW. 1995. *Non verbal predication and head movement*. Doctoral dissertation, MIT

CHOI, HYE-WON. 1996. *Optimizing Structure in Context: scrambling and information structure*. Doctoral dissertation, Stanford University

CHOMSKY, NOAM. 1995. *The Minimalist Program*. MIT Press

CINQUE, GUGLIELMO. 1990. *Types of A-bar dependencies*. MIT Press

CINQUE, GUGLIELMO. 1997. *Adverbs and the Universal Hierarchy of Functional Projections*. To appear in Oxford University Press

COSTA, JOÃO. 1996. Word Order and Constraint Interaction. To appear in *Seminários de Linguística*, Univ. do Algarve, Faro

COSTA, JOÃO. 1996a. Scrambling in European Portuguese. To appear in *Proceedings of SCIL 8*, MIT Working Papers in Linguistics

COSTA, JOÃO. 1996b. Focus in situ: evidence from Portuguese. To appear in *Probus*

COSTA, JOÃO. 1997. Positions for Subjects in European Portuguese. In *Proceedings of the West Coast Conference on Formal Linguistics XV*, CLSI, Stanford

COSTA, JOÃO. Forthcoming. *Word Order Variation. A constraint-based approach.* Doctoral dissertation. HIL/Leiden University

DERBYSHIRE, D. Word Order Universals and the Existence of OVS languages. In *Linguistic Inquiry*, 8

DIESING, MOLLY. 1992. *Indefinites*. MIT press

DIK, SIMON. 1978. *Functional Grammar*. Foris, North-holland, Amsterdam

DUARTE, INÊS. 1987. *A Construção de Topicalização na Gramática do Português.* Doctoral dissertation, University of Lisbon

FASSI-FEHRI, A. 1989. *Agreement, Incorporation, Pleonastics and VSO-SVO order in Arabic*, ms, Faculté des Lettres, Rabat

FROTA, SÓNIA. 1994. Is Focus a phonological category in Portuguese?. in P. Ackema and M. Schoorlemmer (eds). *Proceedings of ConSOLE 1*. HAG

FROTA, SÓNIA. 1997. *Focus and Prosodic Interface.* Paper presented at the Primeiro Encontro Informal de Doutorandos em Linguística, University of Lisbon

GIVÓN, TALMY. 1984. *Syntax: a functional-typological introduction.* John Benjamins Publishing Company

GRIMSHAW, JANE. 1997. Projections, Heads and Optimality. *LI*

GRIMSHAW, JANE AND VIERI SAMEK-LODOVICI. 1995. Optimal Subjects. in *UMOP 18*

GRIMSHAW, JANE AND VIERI SAMEK-LODOVICI. 1996. Optimal Subjects and Subjects Universals. To appear in Barbosa et alii (eds) *Is the Best Good Enough?*, MITWPL

HERNANZ, MARIA LLUISA AND JOSÉ MARIA BRUCART. 1987. *La sintaxis.* Editorial Crítica

HUANG, JAMES. 1984. On the distribution and reference of Empty Pronouns. In *Linguistic Inquiry*

KAYNE, RICHARD. 1994. *The Antisymmetry of Syntax*. MIT Press, Cambridge

KEENAN, Ed. 1976. On Remarkable Subjects in Malagasy. In Li (ed.) (1976)

KOOPMAN, HILDA AND DOMINIQUE SPORTICHE. 1991. Subjects. In *Lingua*

LEGENDRE, GERALDINE, COLIN WILSON, PAUL SMOLENSKY, KRISTIN HOMER AND WILLIAM RAYMOND. 1995. Optimality and Wh-extraction.in *UMOP 18*

LI, CHARLES. 1976. *Subject and Topic*. AcademicPress, New York

MARTINS, ANA MARIA. 1994. *Os Clíticos na História do Português.* Doctoral dissertation, University of Lisbon

MATEUS, MARIA HELENA ET ALII. 1989. *Gramática da Língua Portuguesa*. Editorial Caminho

MCCARTHY, JOHN AND ALAN PRINCE. 1994. *The Emergence of the Unmarked*. Ms, Rutgers University and Umass, Amherst

MCCLOSKEY, JAMES. 1992. *On the Scope of Verb Movement in Irish*. Ms, UCSC

ORDOÑEZ, FRANCISCO AND ESTHELA TREVIÑO. 1995. Los sujetos y objetos preverbales en español. Paper presented at the 5th Colloquium on Generative Grammar, Coruña, Spain

ORDOÑEZ, FRANCISCO. To appear. Postverbal asymmetries in Spanish. In *NLLT*

OUHALLA, JAMAL. 1991. *Functional Categories and Parametrization*. Routledge

PINTO, MANUELA. 1997. *Licensing and Interpretation of Inverted Subjects in Italian*. Doctoral dissertation, Utrecht University

PRINCE, ALAN AND PAUL SMOLENSKY. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. ms, Rutgers and University of Colorado [to appear in MIT Press]

RAPOSO, EDUARDO. 1986. On the Null Object Construction in European Portuguese. In Osvaldo Jaeggli and C. Silva-Corvalán (eds.) *Studies in Romance Linguistics*, Foris: Dordrecht

RAPOSO, EDUARDO. 1996. *Towards a unification of Topic Constructions*. Ms, UCSB

RAPOSO, EDUARDO. 1997. *Definite/Zero Alternations in Portuguese (Towards a unification of Topic Constructions)*, paper presented at LSRL

RIZZI, LUIGI. 1982. *Issues in Italian Syntax*. Foris, Dordrecht

SAMEK-LODOVICI, VIERI. 1996. *Constraints on Subjects: an optimality-theoretic analysis*. Doctoral dissertation, Rutgers University

SOLÀ, JAUME. 1992. *Agreement and Subjects*. Doctoral dissertation, University of Barcelona

SPROAT, R. 1985. Welsh Syntax and VSO structure. In *NLLT*

SZABOLCSI, ANNA. 1981. The Semantics of Topic-Focus Articulation. in *Formal Methods in the Study of Language*

TRAVIS, LISA. *Parameters and Effects of Word-Order Variation*. Doctoral dissertation, MIT

VIKNER, STEN. 1997. *V°-to-I° Movement, DO-insertion, and Negation in Optimality Theory*. paper presented at NIAS, April

ZUBIZARRETA, MARIA LUISA. 1995. Word Order, Prosody and Focus. Ms, USC

ZWART, C. JAN WOUTER. 1993. *Dutch Syntax: a minimalist approach*. Doctoral dissertation, Groningen