The sequences 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 chapter, 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 (OT; Prince and Smolensky 1993), and suggest that the observed variation is 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 chapter is as follows: section 7.1 illustrates the word order patterns examined. Section 7.2 spells out theoretical assumptions, and section 7.3 the constraints used and the ranking proposed for Portuguese. In section 7.4, I derive the other word orders considered by reranking the constraints proposed for Portuguese.

This work is intended 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, 1998), Samek-Lodovici (1996), Costa (1997c, 1998), 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?

7.1 Problem

A well-known fact concerning linguistic variation is that different languages display different basic word orders, where basic word order means 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?” (among other sources, see Li 1976; Dik 1978).
The following examples show different base word orders for different languages. Portuguese is an SVO language, as (1) shows.

(1) Portuguese
   O João comeu o bolo.
   ‘John ate the cake.’

References to the basic word order in Portuguese are made in Amihai 1992, Duarte 1987, and Mateus et al. 1989, among many others.

In other Romance languages, the same basic word order can be found. As Rizzi (1982) and others observe, 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 Hernanz and Brucart (1987).

(3) Spanish
   Juan ha visto a María.
   ‘Juan has seen María.’

It is worth noting that the situation in Spanish is not that simple; authors disagree on what the basic word order in Spanish is. Some studies claim that the basic word order of Spanish is VSO (Ordóñez and Treviño 1995; Zubizarreta 1995).

(4) Spanish
   Comió Juan los guisantes.
   ate Juan the peas

In this chapter, I will not take a position concerning the correctness of either of these two descriptions. 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. I will name the SVO variant Spanish A, and the VSO variant Spanish B.

Greek is another language with basic VSO order, as Alexiadou and Anagnostopoulou (1995) note.

(5) Greek
   Πανδρόφηκε το Πέτρο την Ήλεκτρα
   married the Peter-NOM the Electra-ACC
   ‘Peter married Electra.’

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 Electra
   ‘Peter married Electra.’

   Arabic is 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-un.
   fut-3ms-buy Zayd house
   ‘Zayd will buy a house.’

   Other VSO languages include Berber, Chamorro, and Celtic; see examples (8) to (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-tailta si Maria i lepblu.
   3sg-read the 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 differences between them that have led some authors to analyze them as belonging to distinct typological groups (see Ouhalla 1991; 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. For details, 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 follows both the verb and the object.

(13) Malagasy
    Nahita an-dRabe Rakoto
    saw ACC-Rabe Rakoto
    ‘Rakoto saw Rabe.’

In this chapter, I will not consider languages in which the object precedes both the subject and the verb in the unmarked case. Such languages are rare (see Derbyshire 1977 for OVS in Carib), and debate surrounds their status as neutral in the attested cases (see Givón 1984 for discussion). But I will 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:

1. What does it (formally) mean to be the unmarked word order?
2. Why is there variation at the base?
3. Is the unmarked word order truly basic or is it in turn 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 chapter I will argue, along the lines of McCarthy and Prince (1994), that OT provides such a formalism.

The second question is of a general nature: If all these word orders are felt to be basic in the respective languages, why should there be different basic structures? That is, why is there not a uniform basis? The stand taken in this chapter is that there may be a uniform basis from which these basic word orders are derived. Different orders emerge as unmarked because different languages correspond to different constraint rankings; thus different basic word orders are unmarked with respect to different constraints. Following the reasoning that underlies the answer to the second question, the position taken in this chapter is that the basic word order does not necessarily correspond to an underlying representation. That is, once we assume that the surface base representation may already be the result of compliance with 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 chapter, I will not take a strong position with respect to whether there should be only one base word order (like Kayne does). Nevertheless I would like to stress 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.

7.2 Theoretical Background: Optimality Theory and Emergence of the Unmarked

According to OT, 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. Tableau T7.1 exemplifies a potential OT diagram. This tableau 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 larger and could 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 that are normally invisible because all candidates that 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. McCarthy and Prince (1994) call the possibility of seeing the effects of a constraint that normally is violated the Emergence of the Unmarked. Since the Emergence of the Unmarked is a relevant distinctive feature of OT, most works on syntax and phonology within this framework exploit it. This notion is crucial in this chapter, since I will argue that unmarked or base word orders are just a subcase of the Emergence of the Unmarked. The next section shows this for Portuguese.

Tableau T7.1

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇝ Candidate 1</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Candidate 2</td>
<td>*</td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>
7.3 Deriving the Unmarked SVO Order in Portuguese

In Costa 1997c, 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: The rightmost constituent in a clause is focused.
   b. SUBJCASE: Subjects are Case-licensed in SpecIP.
   c. OBJCASE: Objects are Case-licensed in SpecAgrOP.
   d. STAY: Do not move.
   e. TOPFIRST: Topics are sentence initial (inspired by Li 1976); non-topics cannot be topicalized. Failed by topics that are not sentence initial, and by topicalized non-topics.

These constraints are independently motivated and have been used in different domains, although not necessarily as violable constraints (for ALIGNFOCUS, see Grimshaw and Samek-Lodovici 1995, 1998; for CASE, Chomsky 1995 and Vikner, chap. 14, this volume; for STAY, Grimshaw 1997).

These constraints are in conflict, since STAY penalizes 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 leftward to SpecIP. Similarly, TOPFIRST conflicts with OBJCASE by forcing an object to surface in the topic position instead of in the case-licensing position. In OT, these conflicts are resolved language-internally, in accordance with the ranking of constraints that constitutes the language's grammar.

It may look at first glance as though I am departing from the assumption that different types of movement may feed one another, for example allowing an object to appear in a topicalized position after moving through SpecAgrOP for case purposes. I do not reject the standard view that different constraints may be satisfied through chains. However, the crosslinguistic evidence presented here and in Costa 1998 appears to point to the conclusion that compliance with the constraints above implies that XPs surface in given positions. If derivations were enough to satisfy conflicting constraints, uniform derivations should be found crosslinguistically, and subjects should be able to surface at the rightmost position in the sentence, after passing through SpecIP. Since languages differ in having obligatory or discourse-conditioned subject movement, it seems that the derivational approach may not do the whole job alone (see Costa 1998 for a synthesis of the derivational and OT approaches). For cases in which there is no evidence for the need to satisfy a constraint at the surface, chains may be established. Such a case is the theta criterion. Arguments are generated
VP-internally for the sake of being assigned theta roles, but need not stay there for
the theta criterion to be satisfied.

For Portuguese, I proposed in Costa 1997c that the ranking specified in (15) may
c characterize the behavior of Portuguese word orders.¹

(15) Ranking for Portuguese

\{AlignFocus, TopFirst\} \gg SubjCase \gg Stay \gg ObjCase

This ranking determines the following: Subjects only move to SpecIP if they are not
focused. Objects never move to SpecAgrOP; they stay in their base position if they
are focused and move, adjoinging to VP, if they are not focused. That this is the be-
behavior of arguments in Portuguese has been argued in Costa 1997, 1996b. (For other
approaches to similar word order alternations, see, among other sources, Ambar
1992; Duarte 1987; Barbosa 1995, 1996; Martins 1994. For details concerning the
implementation of the OT approach, see Costa 1997c.) The top ranking of Align-
Focus and TopFirst 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 word order is observed in contexts answering the question “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.
   Maria broke a plate
b. #Partiu a Maria um prato.
c. #Partiu um prato a Maria.
d. #A Maria um prato partiu.
e. #Um prato a Maria partiu
f. #Um prato partiu a Maria
   ‘Maria broke a plate.’

The argument for the emergence of SVO as unmarked goes as follows: from a func-
tional point of view, all these orders are equivalent, provided that the rightmost
constituent of the sentence bears the main stress (see 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
1996 for discussion), the whole sentence may be interpreted as focused independently
of the word order displayed. For hearers to interpret the whole sentence as focused, it
is enough not to hear any high stress before the end of the sentence. When they hear
an SVO sentence with the main prominence on the object, they may interpret it as
focus on the object, on the VP, or on the whole sentence, depending on which con-
stituent is taken as the domain of focus. The same holds for a VSO sentence with
main prominence on the object: hearers may interpret the domain of focus as the object alone, the subject and the object, or the whole sentence.

Similarly, since there is no topic feature at the input for these sentences, $\text{TopFirst}$ is vacuously satisfied by all word orders that do not involve topicalization, since no sentence-initial element will be the topic. Moreover, if some element is left-dislocated, this constraint is violated once, because there is left dislocation of a nontopic element. The crucial point is that, in sentence-focus contexts, $\text{TopFirst}$ is only relevant for ruling out candidates with topicalization of a given constituent.

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 being top ranked, the effects of $\text{AlignFocus}$ and $\text{TopFirst}$ often make the effects of the other constraints invisible. For instance, every time a subject is in focus, it is right-aligned, and consequently $\text{Case}$ has to be violated. If $\text{AlignFocus}$ is not violated by any candidate, as is the case in the sentence-focus context, the effects of $\text{SubjCase}$ become visible. Since this is the next-highest constraint, it will determine the optimal candidate. The winning candidate will be one that does not violate this constraint. This is exemplified in tableau T7.2.

The candidates considered in tableau T7.2 include:

Candidate (a): SVO sentence with subject in SpecIP and object in situ
Candidate (b): SVO sentence with subject in SpecIP and object in SpecAgrOP
Candidate (c): VSO sentence with subject in SpecVP and object in situ
Candidate (d): VOS sentence with subject in SpecVP and object in SpecAgrOP
Candidate (e): OVS sentence with subject in SpecIP and object in SpecCP
Candidate (f): OVS sentence with subject in SpecVP and object in SpecCP

<table>
<thead>
<tr>
<th>Tableau T7.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portuguese: Input: ${V(x,y), \text{Focus} = (V,x,y)}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>a. $[\text{IP} \text{ S} \text{ V} {\text{VP t t O}}]$</td>
</tr>
<tr>
<td>b. $[\text{IP} \text{ S} \text{ V} {\text{AgrOP O} {\text{VP t t t}}}]$</td>
</tr>
<tr>
<td>c. $[\text{IP} \text{ V} {\text{VP S t O}}]$</td>
</tr>
<tr>
<td>d. $[\text{IP} \text{ V} {\text{AgrOP O} {\text{VP S t t}}}]$</td>
</tr>
<tr>
<td>e. $[\text{CP} \text{ O} {\text{IP} \text{ S} \text{ V} {\text{VP t t t}}}]$</td>
</tr>
<tr>
<td>f. $[\text{CP} \text{ O} {\text{IP} \text{ V} {\text{VP S t t}}}]$</td>
</tr>
</tbody>
</table>
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 assume that this alternation should be analyzed at a deeper level than the one proposed here. Since SOV relates to directionality of theta-role assignment (Hoekstra 1984; Travis 1984; among others), and thematic information in OT is partially expressed in the input (see 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 a goal is beyond the scope of this chapter.

Note nevertheless that it is conceivable to incorporate an alternative 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 SpecAgrOP. This would mean compliance with SUBJCase, OBJCase, 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 1997b), where I explain Dutch scrambling resorting to head finality, I will not adopt the latter hypothesis, stressing though that this is not crucial for the present chapter. The set of representations left out are all those in which the verb has not moved to I. The representations are not included for the sake of brevity and clarity: this avoids adding constraints on V-to-I to the tableaux, making their interpretation much easier. For an analysis of V-to-I within OT, see Vikner (chap. 14, this volume). Again, including those candidates would not change the argument developed in this chapter. Since I am not questioning the standard evidence for or against V-to-I movement in the languages under discussion, and since the constraints on V to I movement (see Vikner, chap. 14, this volume) do not conflict with the constraints to be used in the analysis, I will leave correlating the results of Vikner’s work and the results of this chapter for future research.

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 SpecIP, SUBJCase is satisfied. Satisfying OBJCase implies that the Object also moves out of VP, incurring one mark for STAY. 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 SpecIP, violating SUBJCase. This is a case of the Emergence of the Unmarked: in general, violation of SUBJCase is allowed in Portuguese, if the subject is focus. In that case, the effects of SUBJCase are not visible. Here, since ALIGNFocus is vacuously satisfied by all candidates, the effects of SUBJCase become apparent, and representations that do not cause this constraint to be violated are preferred.

Candidate (d) is ungrammatical for the same reason. It fatally violates SUBJCase.

Candidates (e) and (f) are slightly different, since they violate TOPFirst. The reason for the marker on TOPFirst is that there is an unnecessary topicalization of the object.
This is perhaps also penalized by a constraint not represented in the tableau prohibiting the fronting of elements that are not operators (for constraints of this type, see Grimshaw 1997; Samek-Lodovici 1996; Grimshaw and Samek-Lodovici 1995, 1998). It is not unlikely that this constraint may have to be split up into two. There are several conceivable ways to express the same idea that topicalization of nontopics should not be permitted: not only does it violate TopFirst, at defined above, but it also constitutes a violation of *Struc (a constraint penalizing structure; see Grimshaw 1997), since it involves one extra layer of structure. I am calling this constraint *Struc and not OblP (obligatory head) as in Grimshaw 1997, because the focus of this chapter is not heads but phrases in general. The two analyses do not make any different predictions (see Costa 1998 for discussion). In this chapter, I tend to take it for granted that it is excluded because of a constraint that is functional in nature. One advantage of such an approach is that it derives the fact that OS word orders are rare if not nonexistent as unmarked. This follows from the fact that the top-ranked functional/discourse-related constraints immediately filter out such sequences. The only situations where OVS and OSV candidates may be optimal are cases in which there is no movement of the subject, but the object is obligatorily moved to SpecAgrOP. O-initial languages should, according to this line of reasoning, only be possible in languages without V-to-I movement. Since I am not considering V-to-I movement in this chapter, I will leave this prediction unverified. Note that this distinction between TopFirst and *Struc is not a matter of choice between structural and functional approaches, since both constraints are necessary, as will become evident when we look at Celtic. Unless necessary (see the discussion of Celtic below), I will not represent *Struc in the tableaux.

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

### 7.4 Minimal Rerankings, Base Word Orders, and Variation

If the approach advocated here proves valid, it has two important theoretical consequences, since it reinforces two of the premises of OT. First, it confirms that constraints are violable. In fact, the 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, the analysis here achieves another major goal of OT—to provide a formalism to capture linguistic variation—by showing that reranking a universal set of constraints enables us to capture crosslinguistic differences.

### 7.4.1 Portuguese/Spanish A vs. Spanish B/Greek: SVO vs. VSO

The first difference I will examine is the one between SVO and VSO word order of the type described for Spanish by Ordóñez and Treviño (1995) and for Greek by Alexiadou and Anagnostopoulou (1995). As noted, discrepancies exist among speakers of Spanish with respect to SVO versus VSO word order. Explaining these discrepancies in terms of a reranking of several constraints would be undesirable, since that might have consequences for the overall grammar of the language that would be difficult to evaluate; it would also 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 this difference can be explained under a minimal reranking, the null hypothesis is that the same differences in the grammar are responsible for the differences between Portuguese and Greek.

Before presenting the crucial rankings, one must examine the factors that make these languages differ; according to the descriptions in the literature, the differences seem to lie in the behavior of subjects.

According to most descriptions of Portuguese (Raposo 1986, 1996; Mateus et al. 1989; Duarte 1987; Ambar 1992; Martins 1994; Costa 1997), preverbal subjects in Portuguese are in SpecIP (or SpecAgrSP). The same is argued for Spanish by Herranz and Brucart (1987), who also describe Spanish as an SVO language. The argument these authors present for these descriptions is based on the behavior of preverbal subjects as A-moved elements (not intervening for A-bar movement and behaving as A-binders, among other properties).

Conversely, Alexiadou and Anagnostopoulou (1995) for Greek and Ordóñez and Treviño (1995) for Spanish have proposed that these languages are VSO, in the sense that subjects remain in SpecVP and preverbal subjects are merely instances of subject left dislocation. They argue in favor of this analysis on the basis of factors like the following: preverbal subjects either block A-bar movement (Spanish), or interact with wh-phrases in the same way topics do, as (17) suggests for Greek: the interpretation of quantifiers is different depending on their position—only preverbal QPs have a strong reading (see (18)); the issue of construing pronouns as bound variables, which is only possible with postverbal ones, showing that only the latter are in an A-position (exemplified in (19) by data from Catalan, which, according to the authors, behaves like Spanish and Greek).
(17) Greek (from Alexiadou and Anagnostopoulou 1995)
   a. Pjon (*o Petros) ide (o Petros)?
      whom (the Peter-NOM) saw (the-Peter-NOM)
      ‘Whom did Peter see?’
   a’. *Pjos ton Petρo ton ide?
      who the Peter-ACC CI-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 CI-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
      ‘A certain person/one of the people greeted Mary.’
   b. Heretise enas ti Maria. Weak (indefinite) reading
      ‘Someone greeted Mary.’

   a. *Tots els estudiants; es pensen que ells, aprovaran.
      all the students think that they will-pass
      ‘All the students think that they will pass.’
   b. Tots els jugadors, estan convencus que guanyaran ells.
      all the players are convinced that will-win they
      ‘All the players are convinced that they are the ones who will win.’

Notice that, as mentioned earlier, Portuguese subjects are able to co-occur with
fronted constituents and wh-elements (20); asymmetries like the one in (19) do not
exist (see (21)), since 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
      ‘Which books did João read?’
   b. Esses livros, o João leu.
      those books, João read
      ‘Those books, João read.’
On the basis of these asymmetries between the two sets of languages, I conclude that the distinction is finer than what Alexiadou and Anagnostopoulou (1995) propose; they suggest that null-subject languages all instantiate left dislocation whenever a preverbal subject is present. In these languages, it is the verbal agreement of the verb that bears the theta role and is responsible for satisfying EPP features.

In light of these data, it seems to me that null-subject languages do not behave uniformly: some have basic VSO word order and their subjects do not move to SpecIP (Greek and Spanish B); others have basic SVO word order and their subjects may move to SpecIP. The conclusion is then that the crucial difference between these two sets of languages rests on whether preverbal subjects are A-bar or A-bar moved. Such a difference has implications for which of the constraints presented above is satisfied.

Given the definition of the constraints just presented, whenever a subject moves to SpecIP, it satisfies SUBJCase at the expense of violating Stay, since there is one movement operation. In a language in which SUBJCase dominates Stay, it is more important to move the subject than to satisfy economy. This is represented in tableau T7.3. As for left dislocation, I will follow the claim of various authors that this phenomenon does not involve movement but base generation of the left-dislocated element in adjunction to CP (see Duarte 1987; Cinque 1990; Raposo 1996, 1997). As mentioned, when there is left dislocation of the subject, the theta role is assigned to the pronominal agreement (pro or clitics), which never occupies SpecIP. In this chapter, I will follow Duarte’s (1996) claim that left dislocation must be treated as a case of base generation, in accordance with Cinque (1990), among others. Duarte shows that Portuguese differs from other Romance languages in not using left dislocation (the way Spanish, Italian, and Greek do) as a strategy for topic promotion.

The empirical arguments she presents constitute additional reasons for not grouping

<table>
<thead>
<tr>
<th></th>
<th>SUBJCase</th>
<th>Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>![Image of tableau]</td>
<td>![Image of tableau]</td>
</tr>
<tr>
<td>b.</td>
<td>![Image of tableau]</td>
<td>![Image of tableau]</td>
</tr>
</tbody>
</table>
Portuguese with the other null-subject languages. Since Portuguese does not use left
dislocation for promotion of nonsubject topics, it is unlikely to use it for promoting
subjects. If one wants to adhere to the subject-left-dislocation analysis for an SVO
sentence in Portuguese, one has the burden of explaining why left dislocation is
available for subjects only.

Accepting the idea that left dislocation involves base generation of the topic, it is
possible 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
(SpecCP or adjunction to IP, depending on the analysis). That this is indeed the case
is defended by Barbosa (1996), who shows that some preverbal subjects must associ-ate
with subject clitics overtly realized in SpecIP in some Romance languages.

(22) *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 SpecIP 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 SpecIP or base gen-
erating them in topic position is then a matter of whether it is more important to
satisfy STAY (as in (23)) or SUBJCase (as in (24)). The two possible situations are given
in tableaux T7.4 and T7.5, where candidate (a) is the one with subject left dislocation
and candidate (b) has movement of the subject to SpecIP.

(23) STAY ≺ SUBJCase
(24) SUBJCase ≻ STAY

Tableau T7.4 reflects the situation in Spanish B and Greek, where it is more impor-
tant to have an economical representation (without movement) than satisfying CASE.

Tableau T7.4
Greek and Spanish B—Context specification in the input: subject is the topic of
the sentence

<table>
<thead>
<tr>
<th></th>
<th>STAY</th>
<th>SUBJCase</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [CP S [IP pro V [VP t O]]]</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. [IP S V [VP t O]]</td>
<td>**!</td>
<td></td>
</tr>
</tbody>
</table>
Tableau T7.5
Portuguese and Spanish A—Context specification in the input: subject is the topic of the sentence

<table>
<thead>
<tr>
<th></th>
<th>SUBJ CASE</th>
<th>STAY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td><img src="image1.png" alt="image" /></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td><img src="image2.png" alt="image" /></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

In these languages, both candidates violate STAY once because of the movement of the verb. However, the candidate with movement of the Subject to SpecIP involves one more violations of STAY, the second one being fatal.

Tableau T7.5 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 SpecIP.

One may wonder why there is no representation where the subject is base generated in SpecIP, 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 SpecIP would yield a violation of the theta criterion, precluding generation of such candidates. Naturally, this type of analysis raises questions that fall beyond the scope of this chapter—questions related to how the thematic interpretation itself of the left-dislocated subject takes place. In principle, these issues may be addressed 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 (for discussion, see Grimshaw 1997; Legendre et al. 1995; Costa 1998), it does not follow in any way that a subject that does not get its theta role in a local relation with the assigner will not be interpreted as the thematic subject of the verb. Also, if Alexiadou and Anagnostopoulou's analysis is correct, the theta role is assigned to the subject agreement marker, not to the left-dislocated item.

So far, I have just represented the choice between SVO with movement to SpecIP and SVO with left dislocation of the subject. I 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 TopFIRST also plays a role. Consider the following three candidates: SVO with left dislocation, SVO with subject in SpecIP, and VSO. Since the context is sentence focus, all orders will violate TopFIRST once, for the first element of the sen-
tence 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 nontopic element. This is similar to the way I 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 TOPFIRST, it is now possible to see whether the choice is made by SUBICASE 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 SpecIP. This is represented in tableau T7.6. In Greek and Spanish B, it is more important to satisfy STAY than SUBICASE, hence subjects will not move to SpecIP, yielding a more economical representation. This is schematized in tableau T7.7. It is important to note that, in spite of the relevance of TOPFIRST for the exclusion of the candidate with left dislocation of the subject, it is the ranking between STAY and SUBICASE 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 tableau T7.6, SVO is selected, because only in this order has the subject moved to SpecIP, satisfying SUBICASE. In tableau T7.7, because of the reverse ranking, STAY prohibits SVO, since there are two traces in the structure as opposed to one in the VSO order.

**Tableau T7.6**
Portuguese and Spanish A—Context specification in the input: sentence focus

<table>
<thead>
<tr>
<th>Align-focus</th>
<th>Top-first</th>
<th>Subj-case</th>
<th>Stay</th>
<th>Obj-case</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [CP S [IP V [VP t O]]]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. [IP S V [VP t t O]]</td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>c. [IP V [VP S t O]]</td>
<td></td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

**Tableau T7.7**
Greek and Spanish B—Context specification in the input: sentence focus

<table>
<thead>
<tr>
<th>Align-focus</th>
<th>Top-first</th>
<th>Stay</th>
<th>Subj-case</th>
<th>Obj-case</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [CP S [IP V [VP t O]]]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. [IP S V [VP t t O]]</td>
<td></td>
<td>**!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. [IP V [VP S t O]]</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Note that one of the purposes in capturing the difference between SVO and VSO has been achieved: given the difference between Spanish A and Spanish B, which is dialectal, it is not desirable to propose a substantial difference in constraint profiles. Indeed, the difference between the two languages is minimal; only SubjCase and Stay need to be reranked with respect to one another. Statements (25) and (26) indicate the rankings for Portuguese and Spanish A, and Greek and Spanish B, respectively.7

(25) Portuguese and Spanish A
   \{AlignFocus, TopFirst\} \gg SubjCase \gg Stay \gg ObjCase

(26) Greek and Spanish B
   \{AlignFocus, TopFirst\} \gg Stay \gg SubjCase \gg ObjCase

In this way, I can control the consequences of reranking these constraints for other aspects of the grammar of the languages under consideration, since SubjCase has in its scope a quite restricted set of elements. If I were to radically modify the constraint profile, I would be very likely to predict differences between the two dialects of Spanish not observed in actual data.

7.4.2 Postverbal Subjects: VSO vs. VOS

In the preceding section, I 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 (27) illustrates.8

(27) Italian (from Pinto 1997)
   Che cosa è successo?
   What happened?
   a. Beatrice ha scritto lettere d’amore.
      Beatrice has written letters of love
      ‘Beatrice has written love letters.’
   b. Ha scritto lettere d’amore Beatrice.
      has written letters of love Beatrice

One crucial way Italian differs from Portuguese is that though VOS orders are allowed when only the subject is in focus, as in (28), VSO orders legitimate in Portuguese are ungrammatical in Italian, independently of the context (see (29); also noted by Rizzi 1982 and Burzio 1986, among others).
(28) *Italian (from Pinto 1997)*

Chi ha scritto lettere d’amore?

a. #Beatrice ha scritto lettere d’amore.
   Beatrice has written letters of love
   ‘Beatrice has written love letters.’

b. Ha scritto lettere d’amore Beatrice.
   has written letters of love Beatrice

(29) 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 in which VSO orders are legitimate in Portuguese.

(30) Ninguem escreveu nada.

a. Escreveu a Beatriz cartas de amor.
   wrote the Beatriz letters of love
   ‘Beatriz wrote love letters.’

b. ?#Escreveu cartas de amor a Beatriz.
   wrote letters of love the Beatriz
   ‘Beatriz wrote love letters.’

Here I will follow Pinto’s analysis of VOS order. Pinto claims that in spite of the lack of evidence from adverbial placement (see Belletti 1990), VOS orders may be analyzed as an instance of subject in SpecVP with movement of the object to SpecAgrOP, as in the representation in (31).

(31) [P V [ARP O [VP S t]]]

Such an analysis is also defended in Cinque 1999. Cinque argues for this analysis 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, Pinto’s observation that inverted subjects may precede VP-adjuncts (see (32)) confirms this hypothesis, since we would otherwise expect to find right-dislocated subjects following all base adjuncts.

(32) *Italian (from Pinto 1997)*

a. L’arrivato Dante da Firenze.
   is arrived Dante from Florence
   ‘Dante has arrived from Florence.’

b. Ha telefonato Beatrice da Milano.
   has called Beatrice from Milan
   ‘Beatrice has called from Milan.’
The only way to rescue the VP-adjoined position for subjects in sentences like (32) 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 reserved for contrastive foci, on which a restriction of uniqueness applies (see Szabolcsi 1981).

I will therefore follow Pinto (1997) and Cinque (1999) for the structural representation of VOS order in Italian.⁹

The question now is whether we can represent the situation in Italian with the set of constraints used earlier. Before trying possible rankings, it is important to understand where the crucial difference between Italian and Portuguese lies. What seems at stake is that Italian has obligatory object movement to SpecAgrOP, while Portuguese does not have it at all.¹⁰ As the examples above illustrated, not even context may force a VSO word order in Italian. In our system, this means that the constraint driving the movement of the object is more important than constraints related to discourse. That is, the ranking between ObiCase and AlignFocus has to be the one in (33).

(33) ObiCase >> AlignFocus

Since subjects do not behave this way (their position is dependent on the discourse context), the ranking between AlignFocus and SubjCase has to be kept the same as for Portuguese.

(34) AlignFocus >> SubjCase

Since preverbal subjects are in SpecIP, 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 SubjCase (which also explains that scrambling as adjunction to VP is not possible in this language).

(35) SubjCase >> Stay

Since topicalization is possible in Italian (see Rizzi 1995, among others), TopFirst has to be more important than ObiCase.

(36) TopFirst >> ObiCase

We thus arrive at the partial ranking for Italian given in (37).

(37) TopFirst >> ObiCase >> AlignFocus >> SubjCase >> Stay

Though it seems quite different from the ranking for Portuguese presented in (38), notice that again I only had to rerank one constraint (ObiCase). This is highly desirable, since this word order variation takes place within a single language family.

(38) {AlignFocus, TopFirst} >> SubjCase >> Stay >> ObiCase
### Tableau T7.8
Portuguese and Spanish A—Context specification in the input: subject and object are focused

<table>
<thead>
<tr>
<th></th>
<th>ALIGN-FOCUS</th>
<th>TOP-FIRST</th>
<th>SUBJ-CASE</th>
<th>STAY</th>
<th>OBJ-CASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[IP V [VP S t O]]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>[IP V [AgrOP O t [VP S t t]]]</td>
<td>*!</td>
<td></td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>[IP V [vp O [vp S t t]]]</td>
<td>*!</td>
<td></td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>

### Tableau T7.9
Italian—Context specification in the input: subject and object are focused

<table>
<thead>
<tr>
<th></th>
<th>TOP-FIRST</th>
<th>OBJ-CASE</th>
<th>ALIGN-FOCUS</th>
<th>SUBJ-CASE</th>
<th>STAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[IP V [VP S t O]]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>[IP V [AgrOP O t [VP S t t]]]</td>
<td>*!</td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>c.</td>
<td>[IP V [vp O [vp S t t]]]</td>
<td>*!</td>
<td></td>
<td>**</td>
<td></td>
</tr>
</tbody>
</table>

Let us now see how the selection of candidates is done. For this case, it is necessary to consider a context where both subject and object are focused. Tableau T7.8 represents the evaluation tableau for Portuguese, where VSO wins. Tableau T7.9 is the evaluation tableau for Italian, where VOS wins. The VSO candidate and the VOS candidates with movement of the object to SpecAgrOP and with scrambling of the object in conjunction to VP are included in the set of candidates under comparison.

As can be seen in the tableaux, the decision regarding which candidate is optimal is made early in the optimization, 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, again the effects of the Emergence of the Unmarked have to be considered. 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 tableau T7.10, where the candidates listed in (39) are considered.

(39) Candidate (a): SVO with Subject in SpecIP and Object in SpecAgrOP  
Candidate (b): SVO with Subject in SpecIP and Object in situ
### Table 17.10

Italian—Context specification in the input: sentence focus

<table>
<thead>
<tr>
<th></th>
<th>TOP-FRST</th>
<th>ORI-CASE</th>
<th>ALIGN-FOCUS</th>
<th>SURF-CASE</th>
<th>STAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[IP S V [AgrOP O t [VP t t t]]]</td>
<td></td>
<td></td>
<td></td>
<td>*****</td>
</tr>
<tr>
<td>b.</td>
<td>[IP S V [VP t t O]]</td>
<td>*!</td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>c.</td>
<td>[IP S V [VP O [VP t t t]]]</td>
<td>*!</td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>d.</td>
<td>[CP S [IP V [AgrOP O t [VP t t]]]]</td>
<td>*!</td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>e.</td>
<td>[CP S [IP V [VP t O]]]</td>
<td>*!</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>f.</td>
<td>[CP S [IP V [VP O [VP t t t]]]]</td>
<td>*!</td>
<td>*</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>g.</td>
<td>[IP V [VP S t O]]</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>h.</td>
<td>[IP V [AgrOP O t [VP S t t]]]</td>
<td>*!</td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>i.</td>
<td>[IP V [VP O [VP S t t]]]</td>
<td>*!</td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

Candidate (c): SVO with Subject in SpecIP and scrambled Object
Candidate (d): SVO with Subject right dislocated and Object in SpecAgrOP
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 SpecVP and Object in SpecAgrOP
Candidate (i): VOS with Subject in SpecVP and scrambled Object

Let us recap what happens in tableau 17.10. 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-FRST**, since they involve topicalization of a nontopic (similar to what happens in Greek). **OBJ-Case** filters out all candidates with the object in situ or scrambling, precluding the emergence of VSO order. A decision now has to be made between SVO and VOS (candidates (a) and (h), respectively). **ALIGN-FOCUS** does not play any role, hence **SUBJ-Case** will determine the optimal candidate: SVO. The VOS order fatally violates **SUBJ-Case**. Note that in this tableau, it is also possible to observe the crucial ranking between **SUBJ-Case** and **STAY**. If **STAY** would dominate **ALIGN-FOCUS**, the VOS order would emerge, since it only involves three movement operations, against the four movements involved in representation (a).
Summing up, once again, it is possible to describe a language by virtue of a minimal reranking of the assumed constraints and to show that the effects of dominated constraints are visible in sentence-focus context.

7.4.3 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 (40) illustrates.

(40) Malagasy (from Keenan 1976)

\[\text{Nividy mofo ho'an'ny ankizy aho.} \]
\[\text{bought bread for the children} \]
\[\text{I bought bread for the children.}\]

Word order in Malagasy, as opposed to Italian, is rigid: complements of the verb may not follow the subject.

(41) a. *Nividy aho mofo ho'an'ny ankizy.
\[\text{bought I bread for the children} \]

b. *Nividy ho'an'ny ankizy aho mofo.
\[\text{bought for the children I bread} \]

Likewise, SVO orders are not allowed (unless there is some kind of left construction or topicalization that is morphologically marked; see Keenan 1976).

(42) *Aho nividy mofo ho'an'ny ankizy.
\[\text{I bought bread for the children} \]

Evidence for analyzing this word order in the same terms as the analysis proposed for Italian, repeated in (43), comes from the distribution of question, exclamatory, and “no longer” particles. These particles are often used to determine the position of the subject (see, for example, Diesing 1992 for German). In Malagasy, they always precede the subject.

(43) [\[\text{[V} [\text{\Lambda_{TOP}} \text{O} [\text{VP S t t}]])\]

(44) a. Nanome vola an-dRabe ve ianao?
\[\text{gave money acc-Rabe Q you} \]
\[\text{‘Did you give money to Rabe?’} \]

b. Manasa lamba anie Rasoa!
\[\text{washes clothes EXCL Rasoa} \]
\[\text{‘Is Rasoa still washing clothes?’} \]

c. Tsy manasa lamba intsiny Rasoa.
\[\text{not washes clothes longer Rasoa} \]
\[\text{‘Rasoa is no longer washing clothes.’} \]
From the preceding examples, one may hypothesize that subjects in Malagasy are VP-internal.11

Expressing this in terms of ranking, it is necessary to capture the fact that VOS is the unmarked word order. This result may be achieved by using the same ranking as for Italian and changing the relative ranking of \textsc{stay} and \textsc{subjcase}. For Malagasy, the ranking between these two constraints must be the ranking specified by (45).

(45) \textsc{stay} \gg \textsc{subjcase}

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

This, however, cannot be all. The discourse-related constraint \textsc{alignfocus} must be relatively low ranked in this language, since different contexts do not permit different word orders, in contrast to the situation in the other languages considered in this chapter. Since subjects are always sentence final, I propose that \textsc{alignfocus} is ranked below \textsc{subjcase}. I will not make any proposal concerning \textsc{topfirst}; since there are topicalizations in this language, \textsc{topfirst} must be relatively high ranked. However, it is unclear how it interacts with the other constraints. For the sake of exposition, I will leave it top ranked. Additional constraints related to morphologically marked topicalization and focus-related constraints impose requirements on the position of contrastively focused elements. They 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 (tableau T7.11) and see how the ranking proposed derives the word order in Malagasy. Tableau T7.11 shows the irrelevance of \textsc{subjcase} and \textsc{alignfocus} for deriving the unmarked word order of this language. The decision is made by \textsc{topfirst}, which rules out all candidates with left dislocation of the subject; by \textsc{objcase}, which rules out all candidates without movement of the object to SpecAgrOP; and by \textsc{stay}, which decides between candidates (a) and (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 \textsc{alignfocus} dominates everything else and it becomes necessary 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 \textsc{topfirst} (if any; see the discussion above) is, however, the same as the role played by \textsc{alignfocus} in the other languages.

7.4.4 More on VSO: Celtic and Arabic

The preceding discussion on Malagasy VOS as opposed to Italian VOS enables me to give an account of the opposition between VSO in Celtic versus VSO in Arabic, Cha-
Tableau T7.11
Malagasy—Context specification in the input: sentence focus

<table>
<thead>
<tr>
<th></th>
<th>Top-First</th>
<th>Obj-Case</th>
<th>Stay</th>
<th>Sub-Case</th>
<th>Align-Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>IP SV [Agrole O t [VP t t t]]</td>
<td></td>
<td></td>
<td></td>
<td>****</td>
</tr>
<tr>
<td>b.</td>
<td>IP SV [VP t t O]</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>IP SV [VP O t [VP t t t]]</td>
<td>*!</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>[CF S] [IP V [Agrole O t [VP t t]]]</td>
<td>*!</td>
<td></td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>e.</td>
<td>[CF S] [IP V [VP t O]]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>f.</td>
<td>[CF S] [IP V [VP O [VP t t]]]</td>
<td>*!</td>
<td>*</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>g.</td>
<td>IP V [VP S t O]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h.</td>
<td>[IP V [Agrole O t [VP S t t]]]</td>
<td>***</td>
<td></td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>i.</td>
<td>IP V [VP O [VP S t t]]</td>
<td>*!</td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Arabic 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.

Examples (46) to (48), taken from Ouhalla 1991, illustrate these patterns for Berber, Arabic, and Welsh.

(46) 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 ijin teddart.
   Moha fut 3ms buy a house
   '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.'
   Zayd-nom fut (TNS) 3ms (AGR)-buy house-acc
   'Zayd will buy a house.'

(48) 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, will not be integrated here. He notes that Welsh differs from the other languages not only in not having SVO, but also in two other features characteristic of SVO languages: Celtic languages have noninflected infinitives, and the order of inflectional morphemes is TnsAgr. Ouhalla suggests that word order alternations (VSO enables SVO, but not vice versa), the order of morphemes, and the availability of noninflected infinitives are three consequences of one parametric difference: whether, in the clause structure, AGR selects T or T selects Agr. He shows that all the VSO properties follow from a structure where T selects Agr, and the SVO properties from a structure where Agr selects T. Under Ouhalla's analysis, in a language where T selects Agr, the subject ends up in SpecAgrSP, not needing to move further up, though SpecTP remains a legitimate position for topicalization. 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 a language-particular rule stating that subjects are assigned nominative Case in SpecVP.

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, I have been following the claims put forward in the literature for a dialect of Spanish that is VSO, though the language exhibits the order of morphemes TnsAgr, a counterexample to Ouhalla's generalization.13

Let us first consider Arabic. This language has an analysis similar to the one proposed for Greek and Spanish B: VSO in the unmarked case is a consequence of the dominance of SubjCase by Stay. The dependence on discourse context will follow from the high ranking of TOPFIRST and ALIGNFOCUS. Hence, I propose exactly the same analysis as proposed for Greek and Spanish B, represented in tableau T7.12. Departing from the assumption made above for Malagasy that rigid word order is a consequence of the low ranking of the discourse-related constraints, I can now explain why Celtic has a rigid VSO order. This order will emerge in the context of sentence focus, represented in tableau T7.13.
Tableau T7.12
Arabic and Berber—Context specification in the input: sentence focus

<table>
<thead>
<tr>
<th></th>
<th>ALIGN-FOCUS</th>
<th>TOP-FIRST</th>
<th>STAY</th>
<th>SUBI-CASE</th>
<th>ORI-CASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([\text{CP} ; \text{S} ; [\text{IP} ; \text{V} ; [\text{VP} ; t ; \text{O}]])]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. ([\text{IP} ; \text{S} ; \text{V} ; [\text{VP} ; t ; \text{t} ; \text{O}])]</td>
<td>**!</td>
<td>**!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ([\text{IP} ; \text{V} ; [\text{VP} ; \text{S} ; t ; \text{O}])]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Tableau T7.13
Celtic—First attempt

<table>
<thead>
<tr>
<th></th>
<th>STAY</th>
<th>SUBI-CASE</th>
<th>OBJ-CASE</th>
<th>ALIGN-FOCUS</th>
<th>TOP-FIRST</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([\text{CP} ; \text{S} ; [\text{IP} ; \text{V} ; [\text{VP} ; t ; \text{O}]])]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. ([\text{IP} ; \text{S} ; \text{V} ; [\text{VP} ; t ; \text{t} ; \text{O}])]</td>
<td>**!</td>
<td>**!</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. ([\text{IP} ; \text{V} ; [\text{VP} ; \text{S} ; t ; \text{O}])]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that it is still TOPFIRST that decides between candidates (a) and (c). However, the low ranking of TOPFIRST is important to ensure that the same word order will be optimal when the subject is not in focus. This is what happens in tableau T7.14: in this tableau it is necessary to resort to other structural constraints, so that we can see that structural constraints really do determine the VSO word order and that the discourse-related constraints are very low ranked. Hence, in tableau T7.14, I add *STRUC, which prohibits generation of phrase structure and is fatally violated by the candidate with CP projected. If this constraint did not play a role, candidate (a) would be optimal. Given the main claim made in this chapter as well as in Grimshaw and Samek-Lodovici 1995 and Costa 1997a, 1997c that optionality arises when discourse constraints dominate syntactic constraints, while rigid word order arises when the converse occurs, it is not surprising that *STRUC dominates both ALIGNFOCUS and TOPFIRST. If that were not the case, candidate (a) would emerge as optimal, as in Greek and Spanish B. Note that adding *STRUC to the constraint profile is not a stipulation to explain the Celtic facts. This constraint is independently used in the work of Legendre and Bresnan, and the role it plays here confirms the idea that the rigid word order follows from the low rank of ALIGNFOCUS and TOPFIRST.
Emergence of Unmarked Word Order

Table 7.14
Celtic—Context specification in the input: subject is topic

<table>
<thead>
<tr>
<th></th>
<th>Stay</th>
<th>Subj-Case</th>
<th>Obj-Case</th>
<th>*Struc</th>
<th>Align-Focus</th>
<th>Top-First</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([CP S [IP V [VP t O]]])</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ([IP S V [vp t O]])</td>
<td>**</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ([VP S t O])</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

By not adopting the whole of Ouhalla’s analysis, it becomes more difficult to account for 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 SpecVP, while Berber and Chamorro have rich agreement because the verb is in SpecAgrSP. I do not fail to account for the agreement patterns, provided that I adopt the strong lexicalist hypothesis (Chomsky 1995), and as long as I make sure that agreement is explained somewhere else in the model. In fact, morphological alignment constraints may account for agreement (McCarty and Prince 1994; Grimshaw 1997; Legendre, forthcoming). 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, I might lose a syntactic account of agreement, but this way I am able to account for the position of the subjects in a uniform way and without resorting to a language-particular rule. More important, if the analysis of Portuguese subjects proposed above is accepted and if Ouhalla’s suggestion is accurate, one would expect to find differences in agreement depending on whether subjects are in SpecIP or in SpecVP. This prediction is, however, not borne out.

(49) a. Comeram os meninos a sopa.

\(\text{ate-3pl the boys the soup}\)

b. Os meninos comeram a sopa.

\(\text{the boys ate the soup}\)

7.5 Conclusion

In this chapter, I have tried to derive the unmarked word orders of several languages in terms of the Emergence of the Unmarked. I have shown that the effect of constraints normally invisible by virtue of the domination by discourse-related constraints becomes visible in the context of sentence focus. For instance, in languages
where the position of the subject is normally determined by the top-ranked constraintALIGNFOCUS, like Portuguese and Greek, the vacuous satisfaction of this constraint in the context of sentence focus makes the effects of the dominated constraints SUBJ-CASE and STAY visible.

The rankings summarized in (50) were proposed.

(50) a. *Portuguese and Spanish A*
    {TOPFIRST, ALIGNFOCUS} ≫ SUBJ-CASE ≫ STAY ≫ OBJ-CASE

b. *Spanish B, Greek, Arabic, Berber, and Chamorro*
    {TOPFIRST, ALIGNFOCUS} ≫ STAY ≫ SUBJ-CASE ≫ OBJ-CASE

c. *Italian*
    TOPFIRST ≫ OBJ-CASE ≫ ALIGNFOCUS ≫ SUBJ-CASE ≫ STAY

d. *Mologany*
    OBJ-CASE ≫ STAY ≫ SUBJ-CASE ≫ TOPFIRST ≫ ALIGNFOCUS

e. *Celtic*
    STAY ≫ SUBJ-CASE ≫ OBJ-CASE ≫ *STRUC* ≫ {TOPFIRST, ALIGNFOCUS}

The only word order not considered here is SOV, for the reasons discussed in section 7.1 and because accounting for it would mean taking into account the constraints on V-to-I discussed in Grimshaw 1997 and Vikner (chap. 14, this volume). These constraints are beyond the scope of this chapter.

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 expressed by means of minimal re-rankings between constraints.

Notes

Thanks to the following people who commented on preliminary versions of this chapter: Sjef Barbiers, Teun Hoekstra, Géraldine Legendre, Gereon Müller, Yierme Samek-Lodovici, Johan Rooryck, and audiences in Lisbon, Leiden, Utrecht, and Amsterdam.

1. Actually, in Costa 1997c I did not consider TOPFIRST, since this constraint was not relevant there. As will become evident, this constraint is only necessary when OS orders are considered.

2. The discussion here is simplified to make the general argument clearer. 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 in order to make the reading of the diagrams easier, and because I am leaving verb movement out of the discussion entirely.

3. See, however, Barbosa 1995 for a different position, as well as Costa 1997a and Duarte 1996 for counterarguments to her analysis.

4. 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 not considering contexts in which the subject is the focus.
5. 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.

6. This statement must be qualified. I do not have information enabling me to tell whether the difference between VSO and SVO in Spanish is dialectal or ideosectal. For the purposes of this chapter, it is enough to note that variation exists among speakers, and that the differences between the rankings capturing it ought to be minimal. See Legendre 1998 for an account of register variation in French in terms of minimal rerankings of constraints.

7. The rankings proposed also predict that all these languages have scrambling of objects as A-bar movement (in adjunction to VP) (see Costa 1997b), if they have scrambling. This is partially confirmed for Spanish by Ordoñez (forthcoming), who shows that object movement in Spanish is A-bar. In their comparative work on clitic doubling and scrambling, Alexiadou and Anagnostopoulou (1996) show that Greek objects do not move to SpecAgrOP (instead, there is clitic doubling in this language). What these rankings do not predict is that subjects and objects behave uniformly with respect to Case licensing within the same language, as Chomsky (1995) proposes. The facts of Portuguese do not seem to confirm such a proposal, hence I do not see this as a weakness in my analysis.

8. In her description of neutral word orders, Pinto notes, following Calabrese (1991) among others, that some inverted structures are felicitous as answers to the question “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. They are not possible with transitive verbs and are 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.

9. 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).

10. I have not presented the facts here, but there seem to be reasons to believe that Portuguese VOS orders are derived via adjunction to VP instead of movement to SpecAgrOP, a fact explained in Costa 1997c by the domination of OutCase by Stay. The explanation is as follows: by satisfying OutCase, a representation is forced to project AgrOP, 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.

11. One problem with this analysis is how to explain the obligatory movement to AgrOP of nonnominal 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. This special construction is probably part of the universal candidate set for the relevant input, blocking the use of indefinites in the canonical subject position.

12. But see Legendre (chap. 9, this volume) on Breton for an example of a Celtic language that also allows SVO and OVS orders.
13. Note that nothing prevents an extension of Ouhalla’s analysis from Celtic to Spanish: Spanish might be an Agr-initial language with the language-particular rule assigning nominative Case in SpecVP.

14. One possible analysis not involving *STRUC and still deriving the correct 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 tableau T.14 violate *STRUC 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. Gereon Müllner has suggested to me that, if in Celtic there is no option of inserting pro to make sure that the theta role of the subject is discharged, the analysis without *STRUC still makes the right predictions. The VSO candidate is still the winner since it is the one involving fewer movement operations. I cannot adopt this suggestion, since topicalization in Celtic is often associated with strong agreement (see Tallerman 1996 for Welsh), not differing from what was observed for null-subject languages in Barbosa 1995 and Alexiadou and Anagnostopoulou 1995. Tallerman (1996) shows that in Welsh, strong agreement occurs in constructions involving pronominals and topicalization, and not in constructions involving movement. Since topicalization (abnormal sentences in Tallerman 1996) is associated with strong agreement, it must not be considered a case of movement, as Müller’s proposal would suggest. I must therefore resort to *STRUC and cannot explain the Celtic behavior in terms of economy of movement alone.

References


Emergence of Unmarked Word Order


Emergence of Unmarked Word Order


