A Case Study of Prosody Driven Language Change.

From Classical to Modern European Portuguese.

by

A. Galves* and C. Galves**

Abstract:

Prosody plays a crucial role in grammar selection, restricting the possible values of the parameters to be set. In the present work we present a formal account of this claim, using the Thermodynamical Formalism. In this framework, the sample of positive evidence presented to the child is chosen according to a Gibbs state defined by the potential associated to the prosody. Given a sample of sentences of the parental grammar and given a prosodic pattern, the child choses the grammar according to a maximum likelihood principle. We argue that such a model accounts for both language acquisition and language change. As an example we study the change in clitic placement from Classical to modern European Portuguese.

Partially supported by CNPq grants 301301/79 (AG) and 301086/85 (CG). This research is part of FAPESP’s Projeto Temático Transição de Fase Dinâmica em Sistemas Evolutivos.

* Instituto de Matemática e Estatística, Universidade de São Paulo, BP 20570, 01452-970 São Paulo, SP, Brasil, galves@ime.usp.br

** Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, BP 6045, 13081-970, Campinas, SP, Brasil, galvesc@ccvax.unicamp.br
1. Introduction.

An important syntactic change took place in European Portuguese at the beginning of the 19th century, which affected both phrasal order and clitic placement. In this paper we argue that this change was driven by an alteration of the prosodic pattern of words and sentences which took place by the end of the 18th century.

As proposed by Lightfoot (1979), grammatical changes occur in the acquisition process. As a consequence of this assumption, any model of language change must be based on a model of grammar acquisition. In the Principles and Parameters approach to grammar \(^1\), the child acquires his native language by fixing the values of a finite set of parameters defined by Universal Grammar. Since the set of the values each parameter may assume is itself finite, language acquisition is actually selection of an element in a finite set.

In the present paper we introduce a probabilistic model of grammar selection which attributes a leading role to prosody. On one hand, this model describes the way a sample of morphological expressions of sentences is offered to a learning child. On the other hand, it describes the way the learning child selects a grammar based on the sample of positive evidence available to him. In both cases, prosody plays a crucial role.

The probabilistic characteristic of the model mimics the fact that the sample of sentences provided to the learning child as positive evidence is a consequence of successive choices made essentially in a random way, obeying only the restrictions of the parental grammar and prosody. We make the hypothesis that this sample is chosen according to a probability measure depending on both the parental prosodic pattern and on the parental grammar.

The probabilistic characteristic of the model also expresses the complex nature of the identification principle which guides the learning child. Given a sample of sentences, the child chooses a grammar by attributing structures to the morphological expressions present in this sample. S/he accomplishes this task by looking for the structure which fits better to the parental prosodic pattern. We claim that this is done through a procedure which is reminiscent of the Statistical Physics approach to pattern recognition.

The Thermodynamical Formalism provides a suitable framework, in which the notions of prosodic pattern and grammar can be put together in the definition of the probability measure governing the choice of the sample of positive evidence. Roughly speaking, given a discursive context, the grammar says which sentences are available and prosody says what is the probability to choose a sentence among all the available ones.

This model accounts for the robustness of language acquisition even in the presence of a restricted sample of sentences provided as positive evidence. It also accounts for language change. Depending on the prosodic pattern, the identification procedure may lead the learning child to chose a grammar which differs from the parental one. As an example we study the change in clitic placement from Classical to modern European Portuguese.

D. Lightfoot called his presentation book on Generative Grammar *The Language

---

\(^1\) The Principles and Parameters approach has been developed in a systematic way by Chomsky and others from the seminal Pisa Conferences in 1979 till the Minimalist Program (Chomsky 1993, 1994). For a general presentation of this model, we refer the reader to Chomsky (1986).
Lottery. A lottery supposes the existence of a probabilistic device. Our aim in this paper is to explain how it works.

This paper is organized as follows. In section 2 we present the data which are relevant for our description of the change from Classical to modern European Portuguese. A formal description of those facts, based on the Minimalist version of the Principles and Parameters model, is given in section 3. In section 4 we present the identification model which will be the guideline of the whole study. In section 5 we define the probability measures governing the choice of CIP and EP clitic clauses. With all these elements at hand, in section 6, we show how a prosodic modification can be used to explain the jump from CIP to EP. Finally, section 7 is devoted to a general discussion of the model.

2. The data.

We now present the syntactic and phonological data which are relevant for our analysis. First, Portuguese has always been a SVO language, as exemplified in 0.

0) Paulo ama Virgínia.
   Paulo loves Virgínia.

From at least the 16th century until the beginning of the 19th century, in root affirmative sentences with non-quantified subjects, both proclisis and enclisis were possible, as exemplified in 1 and 2.

1) Paulo a ama.
   Paulo her loves.
   “Paulo loves her”.

2) Paulo ama-a.
   Paulo loves-her.
   “Paulo loves her”.

During the 19th century a change affecting the syntax of clitic-placement occurred in the language spoken in Portugal (cf Benincà in press, Salvi, 1990 and Torres Morais, 1995). As a result, sentences like 1 became agrammatical and 2 remained as the only option for root affirmative sentences with non-quantified subjects. This change, however, did not concern sentences like 3 with quantified or Wh-subjects in which proclisis was, and continues to be, the only option.

3) Quem a ama?
   Who her loves?
   “Who loves her?”.

We shall call Classical Portuguese (henceforth CIP) the language generating sentences 1, 2 and 3, and Modern European Portuguese (henceforth EP) the language generating

---

2 At the same time, as shown by Salvi (1990) and Torres Morais (1995), important changes showed up in word order. The frequency of V₂ constructions, with non subjects in first position, fairly high in CIP, not only decreased drastically, but also their occurrence are now restricted to focalized XPs.
The change in clitic placement from ClP to EP can be seen in two tables from Torres Morais (1995). The first table presents data extracted from works by five authors born from the last decade of the 17th century to the second half of the 18th century (the precise references are given at the end of this paper). It shows the typical situation of Classical Portuguese where, in contexts where there is variation, proclisis is clearly dominant. In the sample considered by Torres Moraes, the occurrence of enclisis does not exceed 44% of the total.

Table 1. Clitics in Classical Portuguese

<table>
<thead>
<tr>
<th>Author (birth year)</th>
<th>Proclisis</th>
<th>Enclisis</th>
<th>% Enclisis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gusmão (1695)</td>
<td>27</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Castro (1700)</td>
<td>15</td>
<td>1</td>
<td>7%</td>
</tr>
<tr>
<td>Oliveira (1702)</td>
<td>39</td>
<td>7</td>
<td>16%</td>
</tr>
<tr>
<td>Judeu (1705)</td>
<td>27</td>
<td>6</td>
<td>19%</td>
</tr>
<tr>
<td>Verney (1713)</td>
<td>14</td>
<td>11</td>
<td>44%</td>
</tr>
<tr>
<td>Marquês (1728)</td>
<td>30</td>
<td>10</td>
<td>25%</td>
</tr>
<tr>
<td>Marquesa (1750)</td>
<td>34</td>
<td>23</td>
<td>40%</td>
</tr>
</tbody>
</table>

Table 2 shows quite a different picture. The data presented are from works of writers born between 1799 and 1839. In the first of them, Almeida Garrett, we observe an inversion of the relative frequencies of enclisis and proclisis. It is now enclisis which dominates. In Camilo Castelo Branco and Júlio Dinis, respectively born in 1825 and 1839, proclisis almost disappears. Actually, Torres Moraes notes that the few remaining cases of proclisis in these authors occur in marked contexts. We shall argue below that Almeida Garrett belongs to the last generation of Classical Portuguese speakers, while Camilo Castelo Branco and Júlio Dinis are already Modern European Portuguese speakers.

Table 2. From Classical to Modern European Portuguese

<table>
<thead>
<tr>
<th>Author (birth year)</th>
<th>Proclisis</th>
<th>Enclisis</th>
<th>% Enclisis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garrett (1799)</td>
<td>11</td>
<td>37</td>
<td>77%</td>
</tr>
<tr>
<td>Camilo (1825)</td>
<td>6</td>
<td>70</td>
<td>92%</td>
</tr>
<tr>
<td>Dinis (1839)</td>
<td>3</td>
<td>24</td>
<td>88%</td>
</tr>
</tbody>
</table>

3 The label Classical Portuguese follows the tradition. The term European is meant to distinguish EP from the Brazilian version of Modern Portuguese.

4 An exception is provided by the 17th century texts of Padre Vieira in which, as remarked by Martins (1993), enclisis largely overweights proclisis. This fact leads Martins to conclude that the change from ClP to EP took place during the 17th century. Her conjecture is contradicted by the data reported by Torres Morais. It seems more reasonable to explain this discrepancy by a stylistical choice of Vieira.
This syntactic change was preceded by a phonological one. The prosodic pattern of Portuguese suffered a major modification during the second half of the 18th century. An evidence of such a modification can be found in the reduction of /e/ to /o/, in all the non-stressed syllables. (cf Révah 1954, Silva Neto 1952 and Teyssier 1980). Révah considers this the more serious modification which affected the Portuguese pronunciation since the 16th century because it affects the very structure of the words.

Révah quotes the 19th century Portuguese phoneticist Gonçalves Viana, who shows that the following verse of the 16th century Portuguese poet Camões:

\[E \text{ se vires que pode merecer-te}\]

looses two syllables when pronounced by a 19th actor, because of the elision of the /e/, becoming

\[E \text{ se vir's que pode mer'cer-te.}\]

We shall argue that this prosodic modification is responsible for the syntactic change affecting Portuguese at the beginning of the 19th century.

3. The grammars of clitic-placement in CIP and EP.

The aim of this section is to sketch the grammars involved in the change in clitic-placement from CIP to EP. To do this, let us review the analyses of CIP and of the change to EP proposed by Salvi (1990) and Benincà (in press), as well as the analyses of EP proposed by Madeira (1992), and Manzini (1992).

Madeira (1992) and Manzini (1992) work with the following two hypotheses about clitic placement in EP.

a) Only one functional category contains the clitic and the verb in both proclitic and enclitic constructions.

Hypothesis a, together with Kayne's left-adjunction hypothesis leads to hypothesis b.

b) Proclisis corresponds to a structure in which the clitic has adjoined to the verb in Infl. But in enclitic structures, it is the verb which adjoins to the clitic. In this configuration, the clitic occupies the head of Comp and the verb must raise to Comp in order to bind the clitic, which is an affix.

It is worth noting that this analysis implies that the structure of the enclitic sentences is different from the structure of the sentences without clitics, since the only reason for the
verb to move is the presence of the clitic. Furthermore, Manzini considers the clitic as a kind of operator, entering in complementary distribution with other operators like Wh and Focus. According to her, this is the reason why enclisis is impossible with quantified subjects and in interrogative sentences.

The hypothesis of V-raising to C is also put forth by Salvi (1990) and Benincà (in press), for both CIP and EP. These authors claim that in CIP simple clauses the CP projection is always required. They justify this by the fact that Classical Portuguese, as well as Old Portuguese, and in general Medieval Romance languages, share with the so-called V₂ languages the possibility of the XP V Subject order.

As a consequence, in CIP clauses like 1, they assign the subject NP the Spec/CP position. On the other hand, in CIP enclitic constructions, they claim that the subject NP is adjoined to CP, whose specifier is empty. The crucial idea behind their analysis is that enclisis is forced by the prohibition for the clitic to appear in first position in the clause, id est in CP. If in a sentence like 2, the subject is outside the border of the clause, the only possible position for the clitic is after the verb.

According to them, the reason why sentence 1 becomes impossible in EP is that subjects like Paulo are always outside CP in this language. More precisely, Salvi claims that the difference between CIP and EP is that, in the latter, Spec/CP is no more an available position for non WH and non quantified NP’s. He argues that this explain why at the same time we observe a strong diminution of the order XP V Subj, and its restriction to cases of focalization of the XP.

An alternative account for EP enclitic constructions is proposed by Madeira (1992) and Manzini (1992). According to them, in this language, the subject of 2 is in Spec/CP.

We shall retain the following points from these analyses.

1. Only one functional category contains the clitic and the verb in both proclitic and enclitic constructions. Proclisis corresponds to a structure in which the clitic has adjoined to the verb in Infl.
2. In CIP enclitic constructions the subject lies outside the border of the clause, contrarily to what happens in proclitic constructions.
3. The landing site for the subject in EP enclitic constructions is Spec/CP.
4. The specifier position which is the landing site of non interrogative subjects in CIP is no more available in EP.
5. Enclisis appears in a position entering in complementary distribution with WH and Focus.

We claim that the change from CIP to EP is the result from a reinterpretation of the

---

8 For a discussion of this hypothesis, see Rouveret (1993).
9 This hypothesis is strongly supported by the fact that in CIP the choice between 1 and 2 is only available for NP’s which can be dislocated. This excludes Wh-subjects as quem, as well as quantified subjects like alguém (“somebody”).
10 But the careful reader will remark that from the Minimalist assumptions we present below, it will follow that this specifier position is Spec/AgrS, while Salvi stipulates that it is Spec/Comp.
position of the subject in enclitic constructions. This is what we are going to prove in the remainder of this article.

We shall now address the question of the nature of this change in the framework of the Minimalist version of the Principles and Parameter Theory (Chomsky 1993, 1994). In the present state of the art, there is no straightforward way to formulate the parameter setting corresponding to the five points drawn from the syntactic analyses presented above. We shall suggest an account which seems to correspond to these points, as well as to the spirit if not to the letter of the Minimalist Program.

According to the Minimalist Program, the parameters identifying a particular grammar are the values strong or weak assigned to the features of the functional categories. In this model, the functional categories act as checking points of the computational system. Elements of the structure must move to check their own features. Nominal (NP) features are checked in specifier positions and verbal (V) features are checked in head positions. Moreover, the so-called Greed Principle says that this is the only reason for them to move. A movement occurs in overt syntax if the corresponding checking feature is strong. Otherwise it will occur in covert syntax, after Spell-Out.

Besides checking of morphological properties, Chomsky also considers the possibility of the raising of I to C. He claims that ”V-raising to C is actually I-raising, with V incorporated to I, and is motivated by the properties of the (C,I)-system, not morphological checking of V.” (Chomsky 1993, p.29).

We shall adopt the following four assumptions about the Minimalist Checking Machine.

First, we shall assume that Comp is a potential checking point of several functional features, which exclude each other. This means that just one content for Comp can be selected from the lexicon at each derivation. If two are selected, the derivation crashes.

In both Classical and European Portuguese Comp can host either the usual class of operators, such as Wh and Focus, or an Agr-feature. \(^{11}\) When a Wh-content is selected for Comp, both Wh-phrases like Quem and the verb with the corresponding Wh-features must raise for checking in Comp. Whenever the Agr content is selected, the functional category Agr\(_S\) must raise to Comp at some point of the derivation. \(^{12}\)

We shall say that Agr of Comp is strong if the movement of Agr to Comp takes place before Spell-Out, otherwise we shall call it weak.

Second, we shall assume that when Agr raises to Comp before Spell-Out, then the checking position of the NP-features of Agr\(_S\) is no longer Spec/Agr\(_S\), but Spec/Comp. This means that whenever Comp hosts the Agr content and the grammar sets the value of Agr of Comp as strong, the features of the complex category T-Agr are checked in the Comp checking point. In particular, the subject is assigned Nominative Case in the

\(^{11}\) It is conceivable, even in the Minimalist framework, that the availability of Agr as a content for Comp, in the sense just described, is subject to parametrization. In any case, we claim that if this is a parameter, it was not affected by the change from ClP to EP.

\(^{12}\) This is a tentative formulation of the functioning of what Chomsky calls the (C,I)-system.
Third, we assume that in proclitic constructions the clitic pronoun and the verb are drawn from the lexicon separately and the clitic must adjjoin to the verb during the computation. Whatever property forces the clitic to adjoin to the functional category containing the verb, this movement must take place before Spell-out since clitics have no semantic content (no intrinsic reference) and are invisible to the covert component, i.e. “the subsystem that continues the computation to LF after Spell-Out” (Chomsky 1994). The only position in which this adjunction can take place is AgrS.

Fourth, we assume that the enclitic form is drawn from the lexicon as a single word, with the clitic already suffixed to the verb. This entails that this unit has an extra feature to be checked besides the usual Tense and Agreement features associated to the verb. This feature is checked in Comp, provided that the Agr content has been selected from the lexicon to be hosted by Comp, making it a checking point for clitic features.

With these assumptions, we can now show that the parametric difference between CIP and EP is the value of Agr of Comp. Agr of Comp is weak in CIP, but strong in EP. All the other parameters which concern Infl have the same values in both languages as shown in Table 3. In what follows, Infl describes the features of the complex system T-Agr, and Infl/V and Infl/NP stand for the verbal and nominal features of Infl.

<table>
<thead>
<tr>
<th></th>
<th>Infl/V</th>
<th>Infl/NP</th>
<th>Agr of Comp</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIP</td>
<td>strong</td>
<td>strong</td>
<td>weak</td>
</tr>
<tr>
<td>EP</td>
<td>strong</td>
<td>strong</td>
<td>strong</td>
</tr>
</tbody>
</table>

It is now a simple exercise to verify that this accounts for the analysis of the clauses from CIP and EP given above.

In CIP sentence 1, in virtue of the strong value of the features of Infl, the subject and the verb are immediately dominated by IP (AgrS P), Nominative case is assigned to Spec/AgrS and the clitic is adjoined to AgrS. Since Agr of Comp is weak, AgrS only moves to

---

13 In Bobaljik and Carnie (1992), the incorporation of a functional category, for checking reasons, into the functional category which immediately dominates it, is also assumed to have the effect of making the Specifier of the incorporated head no longer available as a checking position.

14 The same assumption is made by Chomsky (1993) about the overt movement of the auxiliaries in English. For the hypothesis that clitics, and in general weak pronouns have no semantic content, see Cardinaletti and Starke (1993) and Corver and Delfitto (1993).

15 We refer the reader to Benincà and Cinque (1993) for evidence that enclitic forms syntactically behave as single morphological units. By the way, this idea already appears in Mussafia (1886), as pointed out by Benincà (1994).

16 This hypothesis reformulates Madeira’s and Manzini’s analysis of enclisis in the minimalist framework. According to these authors, the verb has to perform an altruistic movement to Comp with the only purpose to serve as a host for the clitic. In our minimalist version, the verb moves to satisfy its own necessities of checking.
Comp after Spell-out. Using I as a shorthand for Infl, the resulting structure for sentence 1 is

1) \([CP \ [IP \ Paulo \ [I \ a \ [I \ ama \ ] ] \cdots \ ]]\).

In sentence 2, the verb must raise to Comp to check the features of the clitic affixed to it. This movement occurs after the checking of the Infl features of the verb, and after Nominative case assignment, since Agr of Comp is weak. Therefore, there is no reason for the subject to move to Spec/Comp. The only position for the subject compatible with the ordering in sentence 2 is outside CP. Therefore the structure of sentence 2 in CIP is

2) \([CP \ Paulo \ [CP \ [C \ ama-a \ ] \ [IP \ \cdots \ ]]\] \).

Last, in sentence 3, the presence of an operator in Comp requires that the verb and the Wh-phrase be respectively in Comp and Spec/Comp. Proclisis is obligatory because enclisis is impossible in the presence of a Wh-content in Comp by assumptions 1 and 4. Therefore the structure of sentence 3 in CIP is

3) \([CP \ Quem \ [C \ [I \ a \ [I \ ama \ ] ] \ [IP \ \cdots \ ]]\] \).

In EP sentence 2, Agr has raised to Comp before Spell-Out because Agr of Comp is strong. Spec/AgrS is not instantiated and the subject is assigned Nominative case in Spec/CP. This configuration licenses enclisis since Comp has an Agr content and can host clitic features. Therefore the structure of sentence 2 in EP is

2) \([CP \ Paulo \ [C \ ama-a \ ] \cdots \ ]\).

As for sentence 3, there is no difference with CIP. The subject and the verb are immediately dominated by CP and enclisis is impossible due to the presence of the operator. Therefore the structure of sentence 3 in EP is

3) \([CP \ Quem \ [C \ [I \ a \ [I \ ama \ ] ] \ [IP \ \cdots \ ]]\] \).

The impossibility of 1 in EP follows from the strong value of Agr of Comp. This forces AgrS to raise to Comp before Spell-Out, and therefore there is no available position for the clitic to adjoin to.

Furthermore, assuming that assumptions 1 to 4 are part of Universal Grammar, the parameter setting for CIP presented above can be unambiguously identified on the basis of the positive evidence provided by the morphological expression of sentences 1-3. This can be done in a straightforward way, using ordering considerations only, as shown below. The procedure has 3 steps.

**First step.** By assumptions 3 and 4, proclisis in sentences 1 and 3 (a ama) and enclisis in sentence 2 (ama-a) shows that there are two available landing sites for the verb, AgrS and Comp. This shows that the Agr-content of Comp is weak, since otherwise AgrS would not be an independent checking point.

**Second step.** The overt presence of the verb in Agr shows that its V-feature is strong.

**Third step.** The presence of the subject NP in preverbal position in 1 shows that the NP-feature of the category in whose specifier Nominative case is assigned is strong. Since the NP (Paulo) has nothing to check in Comp, the principle of Greed says that this category is Infl.\(^{17}\) Therefore the learning child concludes that Infl has strong NP-feature.

---

\(^{17}\) We recall that Infl describes the features of the complex system T-Agr.
Now let us suppose that, for some reason, the learning child receives a sample of positive evidence containing only sentences with enclisis. Then s/he has no way to select her/his grammar on the basis of ordering considerations only. In section 6, we will show that in this situation the learning child assigns the sentence a structure using prosodic considerations.

4. The identification model.

The model has three components: a set of grammars \( \mathcal{G} \), a set \( \mathcal{P} \) of functions defining prosodic patterns and an identification principle.

The notion of grammar we adopt here is the one defined in the Principles and Parameters approach. A grammar is a generative system, obeying the constraints of Universal Grammar, and specified by a finite array of parameters which may be taken as binary without loss of generality. Therefore \( \mathcal{G} \) is just the set of all the possible sequences of values assumed by this array of parameters.

Taking into account the analysis presented in Section 3, the difference between the grammars of clitic-placement in ClP and EP reduces to the value of a single parameter, i.e the one we have called Agr of Comp. Therefore in this case-study, we may reduce \( \mathcal{G} \) to a set with two elements

\[
\mathcal{G} = (G_{\text{ClP}}, G_{\text{EP}}),
\]

where \( G_{\text{ClP}} \) stands for the grammar specified by the array in which Agr of Comp is set to the weak value, and \( G_{\text{EP}} \) stands for the grammar specified by the array in which Agr of Comp is set to the strong value.

Given a grammar \( G \) in \( \mathcal{G} \) let us call \( \mathcal{C}(G) \) the set of the clauses generated by \( G \). Associated to each clause, there is a morphological expression which is the ordered sequence of words entering in the clause, a structural description which expresses the derivation of the clause and which contains in particular the categories which are involved in the clause and, finally, a prosodic contour.

In the present model, we only consider the stress features of the prosodic contour. Let us call stress contour this simplified version of the prosodic contour. The stress contour is based on the metrical grid associated with the syntactic structures of the sentences (cf. Halle and Vergnaud, 1987). As a further simplification, the stress contour will be reduced to the ordered sequence of transitions between clause boundaries, stressed elements and non stressed elements. In particular, we shall not consider the hierarchy between stresses. This will be sufficient to describe the change of the prosodic pattern from ClP to EP.

From now on, let us call stress mark any element of the set \( \{[, %, \sim\} \), where \( [ \) stands for clause boundary, \( % \) stands for stressed element, and \( \sim \) stands for non stressed element.

The prosodic pattern will be defined through a function which assigns a positive real number, i.e., a weight to each stress contour. The reasons of such a definition will appear in a few lines, as soon as we define the probability measure governing the choice of the sample.

For the purpose of the present study, it will be sufficient to consider the functions of the stress contour which are defined in a markovian way, i.e. as the product of the weights of the transitions between successive stress marks. This seems to be a reasonable
hypothesis from the phonological point of view. From the mathematical point of view this hypothesis is not a necessary one but it simplifies the presentation of what follows.

Let us define \( \mathcal{P} \) as the set of functions which associate a real positive number to each ordered couple of stress marks. Let \( p \) be an element of \( \mathcal{P} \). Each stress contour is assigned a weight by the prosodic pattern defined by \( p \). This weight is given by the product of the values of \( p \) in the successive ordered couples of nearest neighbour elements of the contour. For example, if the stress contour is \((s_0, s_1, s_2, \cdots, s_k)\), its weight will be \( p(s_0, s_1)p(s_1, s_2) \cdots p(s_{k-1}, s_k) \).

From now on, we shall call prosodic potential any function \( p \) belonging to \( \mathcal{P} \). \(^{18}\)

For each grammar \( G \) in \( \mathcal{G} \) and each potential \( p \) in \( \mathcal{P} \), there is a canonical probability measure having \( \mathcal{C}(G) \) as sample space. \(^{19}\) Let us call \( \mu_p^G \) this probability measure. The

---

\(^{18}\) In the Thermodynamical Formalism what is usually called potential is \( \log p \).

\(^{19}\) We recall the standard mathematical definition of probability measure on a countable sample space. Let \( \Omega \) be a finite set. A probability measure on \( \Omega \) is any function \( \nu \) which maps subsets of \( \Omega \) on the interval \([0,1]\) and which satisfies the following conditions:

i) \( \nu(\Omega) = 1 \) and,

ii) for any pair \( A_1, A_2 \) of disjoint subsets of \( \Omega \),

\[ \nu(A_1 \cup A_2) = \nu(A_1) + \nu(A_2). \]

In the standard terminology, the pair \((\Omega, \nu)\) is called a probability space having \( \Omega \) as sample space and \( \nu \) as probability measure. Subsets of \( \Omega \) are called events.

Condition i) above says that the probability of the total event is 1. Condition ii) says that the probability of occurrence of at least one among two disjoint events is equal to the sum of the probability of these events. We remark that the probability measure \( \nu \) is completely determined by the values it assigns to the unitary subsets of \( \Omega \), since it follows from condition ii) that the probability of any event \( A \) is equal to the sum of the probabilities of the events obtained by individually taking every element of \( A \).

Informally speaking, the sample space is the "list" of the possible results of a given experiment. For instance, if we play a dice twice successively and we are interested in the precise result of the first and the second trial, we should take \( \Omega = \{(x,y) : x = 1,2,\ldots,6, y = 1,2,\ldots,6\} \) i.e.as the set of all the ordered couples of numbers taking the values 1,2,...,6. However, if we are only interested in the sum of the results of the first and second trial, we could take \( \Omega = \{2,3,\ldots,12\} \). Let us suppose that the dice we use is not biased, which means that all the faces appear with the same probability. In the first experiment, the natural probability measure to consider is the one which gives the same weight to every element of \( \Omega \), i.e. \( \nu(\omega) = \frac{1}{36} \). In the second case, the probability measure does not assign the same value to the different results since it is more likely to have the sum equal to 7, which, in the sample space of the first experiment corresponds to the set of elementary events \( \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \) than to have the sum equal to 2, which is obtained only if the dice falls in face 1 twice. This way, using the probability space described in the first experiment, we can easily see that in the second experiment, the probability measure should be defined as follows \( \mu(\{(2)\}) = \mu(\{(12)\}) = \frac{1}{36}, \mu(\{(3)\}) = \mu(\{(11)\}) = \frac{2}{36}, \mu(\{(4)\}) = \mu(\{(10)\}) = \frac{3}{36}, \mu(\{(5)\}) = \mu(\{(9)\}) = \frac{4}{36}, \mu(\{(6)\}) = \mu(\{(8)\}) = \frac{5}{36}, \mu(\{(7)\}) = \frac{6}{36}. \)
measure $\mu^G_p$ gives to each clause in $\mathcal{C}(G)$ a probability which is proportional to the weight associated by the potential $p$ to the stress contour of the clause. This probability measure is the very kernel of the language lottery. This is the law which governs the choice of each one of the sentences $\eta_1, \eta_2, ..., \eta_n$ which are successively and independently offered to the child as positive evidence.

Now the acquisition process can be defined in the following way. Let $p$ and $G$ be the parental prosodic pattern and the parental grammar. A sample of positive evidence is defined by choosing successively $n$ clauses in $\mathcal{C}(G)$, independently and with probability $\mu^G_p$. Let $S_n$ be the set of the morphological expressions of the sample.

The number $n$ is biologically given. It is the number of choices of clauses entering in the constitution of the sample offered to the child during her/his learning process.

The objective of the learner is to select a grammar in $\mathcal{G}$ on the basis of the knowledge of
i) a prosodic pattern $p$, previously learned;
ii) $G$ and the structure of the probability measure governing the choice of the sample of clauses, both neurologically defined;
iii) $S_n$.

We emphasize that the learning child only receives the morphological expressions of the sampled clauses. S/he must attribute structures to them. The prosodic pattern $p$

For an illuminating introduction to Probability Theory, we refer the reader to the classical book by William Feller (1957).

20 In the standard probabilistic language the clauses are independent and identically distributed random variables, taking values in the set $\mathcal{C}(G)$, each one having law $\mu^G_p$. Independence means that the probability of successively choosing sentences $\eta_1, \eta_2, ..., \eta_n$ is equal to the product of the probabilities of each individual choice, i.e. $\mu^G_p(\eta_1)\mu^G_p(\eta_2) \cdot \mu^G_p(\eta_n)$. Some version of the independence property is needed to prove the Law of Large Numbers which says that the frequency of occurrences of a given event in a sequence of independent and identically distributed trials converges to the probability of the event, as the number of trials increases. This is the basis of any statistical measurement in the classical, non-bayesian, approach. Though sentences are obviously dependent from a discursive point of view, it seems reasonable to suppose that from a syntactic point of view each degree-O sentence is independently drawn from $\mathcal{C}(G)$. As a matter of fact, if instead of assuming independence we assumed some reasonable version of weak dependence, our model would work as well, but the computation would become a little bit more tedious. Incidentally, as we were preparing the revised version of the paper, we became aware of Gibson and Wexler (1994) and Berwick and Nyogi (1993), in which the same assumption of independence is made.

21 The fact that linguistic acquisition takes place in a constant amount of time, independently of the parental grammar, is one of the basic argument for the inateness hypothesis which underlies the Theory of Generative Grammar from the very beginning (Cf, for instance, Chomsky 1975). Since we are assuming that the process of selection of the sentences offered as positive evidence is stationary, the fact that acquisition takes place in a constant amount of time implies that it must be performed with a sample produced by a fixed number of choices.
tells her/him how likely a boundary is to occur between two morphologically expressed elements of the sentence. The learner takes advantage of this hint by using the Maximum Likelihood Criterion as his/her identification principle. 22 According to this principle, the grammar selected will be the one which maximizes the probability of the choice of $S_n$.

The instantiation of this model in the case of the grammars of clitic-placement in ClP and EP will be done in the next section.

5. The language lottery.

To construct the probability measures associated with ClP and EP we first need to write the stress contours of clauses 1, 2 and 3 in ClP and 2 and 3 in EP.

Given a bracketed clause, we assign a stress mark $'$ to each stressed word, a stress mark $\sim$ to each unstressed word and a stress mark $[\ ]$ to the CP boundaries 23. We remark that the only non stressed word we are considering is the clitic pronoun when it appears in preverbal position (proclisis). We recall that in the enclitic form the verb and the clitic constitute a single stressed word.

This defines a map from the set of the clauses of a language into the set of finite ordered sequences of stress marks. Let us call $F$ this map. The stress contour of a clause is its image by $F$. The function $F$ codes the ClP clauses 1, 2 and 3 as follows, where the stress contour is indicated in the upper line and the bracketed clause in the bottom line.

1) $[ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ ] [CP [IP Paulo [I a [I ama ] ] ] ] \ ]$.

2) $[ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ ] [CP Paulo [CP [C ama-a ] [IP $\cdots$ ] ] ] ]$.

3) $[ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ ] [CP Quem [C [I a [I ama ] ] ] [IP $\cdots$ ] ] ]$.

The function $F$ codes the EP clauses 2 and 3 as follows.

---

22 The classical method of maximum likelihood estimation is one of the main tools in Statistics. It was promoted by R.A. Fisher in his classical 1925 paper. It may be informally described as follows. Let us suppose we have two boxes containing white and black balls. The first box has 999 black balls and only 1 white ball. The second box has 999 white balls and only 1 black ball. A sample of one ball is drawn from one of the boxes. We know the composition of each box, but we ignore which of them was used to produce the sample. The statistician’s task is exactly to find out which box was used, given the sample. Let us suppose that a black ball was drawn. The probability of such a result with the first box is $\frac{999}{1000}$. The probability of such a result with the second box is $\frac{1}{1000}$. Therefore, it seems reasonable to guess that the sample was drawn from the first box, since the probability of the result with this box is larger than with the other. This is precisely the content of the Maximum Likelihood Criterion. For an elementary example of maximum likelihood estimate we refer the reader to Feller (1957), section II.6.

23 The assignment of the stress marks $'$ and $\sim$ is nothing but a simplified version of the metrical grid. This is sufficient to describe the evolution from ClP to EP.
From now on, to simplify the notation we shall indicate ClP sentences 1, 2 and 3 as 1\text{ClP}, 2\text{ClP}, and 3\text{ClP} respectively, and EP sentences 2 and 3 as 2\text{EP} and 3\text{EP} respectively.

Let \( p \) be an element of the set of prosodic potentials \( \mathcal{P} \). The probability measures we consider give to each clause a probability which is proportional to the weight associated to its stress contour by \( p \). Therefore the probability measure \( \mu^\text{ClP}_p \) defined by the potential \( p \) on \( \mathcal{C}(G_{\text{ClP}}) \) is given by

\[
\mu^\text{ClP}_p(1_{\text{ClP}}) = \frac{p([], t)p(t, \sim)p(\sim, t)}{Z(G_{\text{ClP}}, p)}
\]

\[
\mu^\text{ClP}_p(2_{\text{ClP}}) = \frac{p([], t)^2p(t, \mid)}{Z(G_{\text{ClP}}, p)}
\]

and

\[
\mu^\text{ClP}_p(3_{\text{ClP}}) = \frac{p([], t)p(t, \sim)p(\sim, t)}{Z(G_{\text{ClP}}, p)}
\]

where

\[
Z(G_{\text{ClP}}, p) = 2p([], t)p(t, \sim)p(\sim, t) + p([], t)^2p(t, \mid)
\]

is the normalization factor which makes \( \mu^\text{ClP}_p \) a probability measure, id est \( Z(G_{\text{ClP}}, p) \) is the sum of the values attributed by the prosodic pattern to the stress contours of the sentences belonging to \( \mathcal{C}(G_{\text{ClP}}) \).

We remark that since clauses 1 and 3 have the same stress contour, their probabilities are the same.

The probability measure \( \mu^\text{EP}_p \) defined by the potential \( p \) on \( \mathcal{C}(G_{\text{EP}}) \) is given by

\[
\mu^\text{EP}_p(2_{\text{EP}}) = \frac{p([], t)p(t, \mid)}{Z(G_{\text{EP}}, p)}
\]

\[
\mu^\text{EP}_p(3_{\text{EP}}) = \frac{p([], t)p(t, \sim)p(\sim, t)}{Z(G_{\text{EP}}, p)}
\]

where

\[
Z(G_{\text{EP}}, p) = p([], t)p(t, \mid) + p([], t)p(t, \sim)p(\sim, t)
\]
6. Critical points and change. From ClP to EP.

Having introduced all the elements which appear in the model, we can now explain how a change in the prosodic potential triggered the jump from ClP to EP.

Let us suppose that the parental grammar is $G_{ClP}$ and the parental prosodic pattern is given by $p$. A sample of morphological expressions $S_n$ is offered to the learning child. If the morphological expression of clause 1 is in $S_n$, the child identifies the parental grammar as being $G_{ClP}$, since no clause in $C(G_{EP})$ has the morphological expression of clause 1.

In order to find out if 1 appears in $S_n$, we must compare $n$, which is the number of choices entering in the constitution of the sample, with the mean number of trials which must be performed before 1 appears. Since $S_n$ is obtained by choosing $n$ clauses independently and with probability $p_{G_{ClP}}$, the number of trials before ClP clause 1 appears for the first time has a geometric distribution with mean value equal to the inverse of the probability of the clause. Therefore, an elementary computation whose details are given in a footnote allows us to rewrite this mean value as

$$\frac{1}{1 - p_{G_{ClP}}(1_{ClP})} = 2 + \frac{p([t,])p([t, \sim])}{p(t, \sim)p(\sim, t)}.$$ 

Observe that the ratio

$$\frac{p([t,])p([t, \sim])}{p(t, \sim)p(\sim, t)}$$

gives the relative weight of the transitions stress → boundary, boundary → stress, with respect to the transitions stress → non-stress, non-stress → stress. This shows that the occurrence of 1_{ClP} in $S_n$ depends on the specific features of the parental prosodic pattern expressed in this ratio.

If the mean value $\frac{1}{1 - p_{G_{ClP}}(1_{ClP})}$ is much smaller than $n$ (i.e., if the ratio $\frac{p([t,])p([t, \sim])}{p(t, \sim)p(\sim, t)}$ is much smaller than $n - 2$), then, with very high probability, the morphological expression

$$\eta_{1}, \eta_{2}, \ldots, \eta_{n}$$

be the clauses which are successively chosen as evidence to constitute $S_n$. Let us call $K$ the index of the choice giving 1_{ClP} for the first time. By definition, $\mathbb{P}\{K = k\} = \mathbb{P}\{\eta_{1} \neq 1_{ClP}, \eta_{2} \neq 1_{ClP}, \ldots, \eta_{k-1} \neq 1_{ClP}, \eta_{k} = 1_{ClP}\}$. Let us call $q_{k}$ this probability. Since the random variables $\eta_{1}, \eta_{2}, \ldots, \eta_{n}$ are independent and identically distributed, with law $p_{G_{ClP}}$, $q_{k} = (1 - \mu_{p_{G_{ClP}}(1_{ClP})})^{k-1} \mu_{p_{G_{ClP}}(1_{ClP})}$. This is precisely the geometrical distribution with parameter $\mu_{p_{G_{ClP}}(1_{ClP})}$.

To compute the mean value of the random variable $K$, it is enough to compute the series $\sum_{k=1}^{\infty} kq_{k} = \frac{1}{\mu_{p_{G_{ClP}}(1_{ClP})}}$

24 Let $\eta_{1}, \eta_{2}, \ldots, \eta_{n}$ be the clauses which are successively chosen as evidence to constitute $S_n$. Let us call $K$ the index of the choice giving 1_{ClP} for the first time. By definition, $\mathbb{P}\{K = k\} = \mathbb{P}\{\eta_{1} \neq 1_{ClP}, \eta_{2} \neq 1_{ClP}, \ldots, \eta_{k-1} \neq 1_{ClP}, \eta_{k} = 1_{ClP}\}$. Let us call $q_{k}$ this probability. Since the random variables $\eta_{1}, \eta_{2}, \ldots, \eta_{n}$ are independent and identically distributed, with law $p_{G_{ClP}}$, $q_{k} = (1 - \mu_{p_{G_{ClP}}(1_{ClP})})^{k-1} \mu_{p_{G_{ClP}}(1_{ClP})}$. This is precisely the geometrical distribution with parameter $\mu_{p_{G_{ClP}}(1_{ClP})}$. To compute the mean value of the random variable $K$, it is enough to compute the series $\sum_{k=1}^{\infty} kq_{k} = \frac{1}{\mu_{p_{G_{ClP}}(1_{ClP})}}$.

25 The intermediate steps are the following

$$\frac{1}{\mu_{p_{G_{ClP}}(1_{ClP})}} = \frac{Z(G_{ClP}, p)}{p([t,])p(t, \sim)p(\sim, t)} = \frac{2p([t,])p(t, \sim)p(\sim, t) + p([t, \sim])^2p([t,])}{p([t,])p(t, \sim)p(\sim, t)} =$$

$$\frac{2p([t,])p(t, \sim)p(\sim, t)}{p([t,])p(t, \sim)p(\sim, t)} + \frac{p([t, \sim])^2p([t,])}{p([t,])p(t, \sim)p(\sim, t)} = 2 + \frac{p([t,])p([t,])}{p(t, \sim)p(\sim, t)}$$

15
of clause 1 is in $S_n$, and the learning child identifies the parental grammar as being $G_{ClP}$. This seems to be the situation of Portuguese from the 16th to the 18th century.

Now let us suppose that the prosodic pattern of Portuguese speakers change in such a way that the mean value $\frac{1}{n} \mu_{G_{ClP}}(1_{ClP})$ becomes much bigger than $n$. This characterizes a first critical point.

In this case, with very high probability, $S_n$ will only contain the morphological expression of clause 2. Since both $G_{ClP}$ and $G_{EP}$ produce clauses with this morphological expression, the learning child must now use the Maximum Likelihood Criterion to decide among the two competing grammars. To do this, s/he compares the probabilities $\mu_{G_{ClP}}(2_{ClP})$ and $\mu_{G_{EP}}(2_{EP})$ and choses the grammar associated to the higher probability.

A straightforward and elementary computation shows that

$$\mu_{G_{ClP}}(2_{ClP}) > \mu_{G_{EP}}(2_{EP}) \text{ if and only if } \frac{p(t, |) p(|, t)}{p(t, t)} > 2.$$ 

Observe that the ratio

$$\frac{p(t, |) p(|, t)}{p(t, t)}$$

also expresses features of the parental prosodic pattern. These features are specifically related to the degree of perceptibility of the pause produced by the parental grammar.

A second critical point appears through this ratio. If its value is greater than 2, then

$$\mu_{G_{ClP}}(2_{ClP}) > \mu_{G_{EP}}(2_{EP})$$

and the learning child choses $G_{ClP}$, in spite of the poverty of the sample of positive evidence to which s/he is exposed.

But if this ratio is smaller than 2, then

$$\mu_{G_{ClP}}(2_{ClP}) < \mu_{G_{EP}}(2_{EP})$$

and the learning child becomes a speaker of EP.

The historical data presented in section 3, provides evidence of the reality of this theoretical scenario.

Proclisis and enclisis coexisted in Portuguese until the beginning of the 19th century, attesting the stability of $G_{ClP}$ as the selected grammar. Let us call $\gamma$ the ratio between the frequencies of occurrences of $2_{ClP}$ and $1_{ClP}$. In the model this ratio is given by 26

$$\gamma = \frac{p(t, |) p(|, t)}{p(t, -) p(-, t)}$$

Let us remark that $\gamma$ is precisely the ratio entering in the definition of the first critical point. This allows us to use the historical data given in Tables 1 and 2 to identify the moment at which the first critical point is reached.

26 This follows from the Law of Large Numbers supposing that the number of trials is large enough.
The data reported in Table 1 provide evidence that until the end of the 18th century, proclisis typically overweights enclisis. A more precise estimation of the value of $\gamma$ would require a deeper statistical study, however it is clear that $\gamma$ is smaller than 1. Actually, in Table 1 $\gamma$ ranges from 0, in Gusmão to 0.8 in Verney.

This amounts to saying that the weight of the transitions

\[ \text{stress} \rightarrow \text{non-stress}, \text{non-stress} \rightarrow \text{stress} \]

is greater than the weight of the transitions

\[ \text{stress} \rightarrow \text{boundary}, \text{boundary} \rightarrow \text{stress}. \]

This makes reasonable to suppose that until the end of the 18th century the first critical point had not been reached yet.

At the beginning of the 19th century, the texts of Almeida Garrett indicate that even if proclisis and enclisis still coexist, the value of $\gamma$ has dramatically changed, getting greater than 1. The precise value obtained in Table 2 for Garrett’s sample is $\gamma = 3.3$. This supports the hypothesis that the first critical point had already been reached at Garrett’s generation. Garrett, born in 1799, is still a CIP speaker. One generation later, Camilo Castelo Branco, born in 1825, is already a EP speaker. The second critical point has been reached at some point of this interval.

The change in the ratios $\gamma$ and $\frac{p(t, l)p(l, j)}{p(t, j)}$ can be related to the phonological modifications which occurred in Portuguese, during the second half of the 18th century. The reduction of non-stressed vowels can be interpreted as the tendency to directly drop from a stressed to another stressed vowel, understood in our model as the diminution of the weight given to transitions of the type

\[ \text{stress} \rightarrow \text{non-stress}, \text{non-stress} \rightarrow \text{stress}. \]

At the level of the sentence a first consequence of this is the decreasing of the proportion of proclitic constructions as exemplified above.

The decreasing of the second ratio $\frac{p(t, l)p(l, j)}{p(t, j)}$ corresponds to a second stage in which the tendency referred to above becomes so strong that the transition $\text{stress} \rightarrow \text{stress}$ is given an overwhelming weight with respect to all the other transitions. As a consequence of this, the second clause boundary in an adjunction structure like CIP clause 2 becomes hardly audible. This leads to the selection of the new grammar $G_{EP}$. 

17
7. Discussion.

The two main characteristics of our model are its probabilistic framework and the leading role it attributes to prosody.

The probabilistic approach has its origin in Statistical Physics and has been widely used in different fields of pattern recognition (cf. Geman 1990 for a survey). Modeling the identification procedure of linguistic acquisition as a statistical inference makes it possible to use classical statistical criteria like the Maximum Likelihood Principle. The statistical inference is based on the hypothesis that the learning child knows what is the structure of the probability measure governing the choice of the sample of positive evidence to which he is exposed. This probability measure is defined by prosody, which is previously acquired by the learner, and by the parental grammar, which is not known by the learner and is precisely the parameter to be estimated.

Placing the problem in a probabilistic framework avoids using an extra condition like the Subset Principle. More generally, the problem of understanding how positive evidence is used in the acquisition process can be formulated in a more satisfactory way. In this framework, the ambiguity problem pointed out by Gold (1967) disappears. There is no need, therefore, to invoke conditions like the Subset Principle, to explain acquisition.

The Subset Principle was introduced in Berwick (1985) as a restriction governing the acquisition process. The mathematical basis of the Principle was Angluin’s solution to the Identification Problem raised by Gold (1967), which can be formulated in the following way. Given a family of grammars, how to identify one of them using an increasing sequence of positive evidence only? In Gold (1967) the issue is addressed in a rigorous mathematical way for the first time, through the identification in the limit model. He showed that any class of formal languages over a fixed alphabet, which contains every finite language together with at least one infinite language, cannot be correctly inferred from positive data, for the simple reason that one cannot dismiss any language which contains the target language.

To overcome this difficulty, Angluin (1980) introduces a condition characterizing the families of nonrecursive languages which can be identified from positive data. Informally this condition requires that for any language \( L \) in the class, there exists a distinctive finite subset \( S \) contained in \( L \), such that no language of the family that also contains \( S \) is a proper subset of \( L \). We refer the reader to Angluin (1980) for the mathematical details.

Berwick (1985) calls this condition the Subset Principle and uses it to model up the natural language acquisition procedure. He claims that this principle implies that at a

\[\text{In this paper, we adopt a classical frequentist, non-bayesian, approach. However, as it was pointed out to us by an anonymous referee, it is tempting to consider the possibility of using the bayesian framework to model up the identification principle oriented by prosody as an a priori distribution. This would mean introducing a probability measure, depending on the prosodic pattern } p, \text{ and defined on the set of grammars } G. \text{ Then we would construct an a posteriori distribution using informations provided by the sample } S_n. \text{ Finally we would use a criterion like the Maximum A-posteriori Likelihood Principle (cf. Geman 1990) to select a grammar. This point of view is developed in a work in progress by Galves, Branco and Zuazola.}\]
given stage of the acquisition, the learning child should select the "narrowest possible language consistent with evidence seen so far" (p. 237). The transposition of Angluin’s condition to the setting of natural language acquisition is problematic. Acquiring a natural language means attributing structural descriptions to the morphological expressions contained in a sample of linguistic utterances. As a consequence, it is not clear what picking up the “narrowest possible language” means. For example, in the case of ClP and EP the structures underlying the morphological expression of clauses 2 and 3 are different and the language generated by $G_{EP}$, understood as the set of structural descriptions generated by the grammar, is not contained in the language generated by $G_{ClP}$. In general, a model of language change depending on the Subset Principle will imply in one way or another that changes only occur from “larger” to “smaller” grammars, whatever order is assumed in the class of natural grammars. Such a drift towards simplification contradicts all it is known today about the time evolution of complex systems (For a more extensive criticism of the Subset Principle, we refer the reader to Joshi 1994).

Recently, several papers have made use of probabilistic notions to model language acquisition. Clark and Roberts (1993) model of language change is based on the Genetic Algorithm and therefore shares with ours the use of a criterion of maximum probability to select a grammar. However, they do not take full advantage of this and still invoke the Subset Principle, which is integrated in the fitness metric.

A more decisive step to introducing probabilistic notions in this field has been taken by Gibson and Wexler (1994). They model language acquisition as a Markov chain taking values in the set of grammars defined by the Universal Grammar. This paper changes in the radical way the landscape of the theory as, for the first time, it provides a theoretical framework in which questions of convergence of the learning algorithm can be addressed. However, in this paper, the crucial question of the nature of the identification principle which underlies the acquisition process is not satisfactorily formulated. We claim that the identification principle cannot be reduced to a yes/no question about whether a grammar generates a sentence of not. Our study of the change from ClP to EP shows that the generation who performed the change received positive evidence which could be produced by both grammars, and had to make a decision choosing the one which assigned these sentences the structure adjusting better to a given prosodic pattern. As a matter of fact, Gibson and Wexler’s algorithm can be generalized in such a way that prosodic and structural considerations can be taken into account. This is developed in Cassandro, Galves and Galves (in progress). We also refer to Nyogi and Berwick (1993) for a mathematical presentation of the Triggering Learning Algorithm, and finally to Frank and Kapur (1994) for a very interesting discussion of the notion of trigger in the context of the GW model.

Our approach addresses the question of the place occupied by prosody in the Principles and Parameters model. In the present state of the theory, the Phonological Form is an output of the computational system (cf. Cinque 1993). In the Minimalist version of the theory, the levels of representation $PF$ and $LF$ are instructions provided by the grammar for the Articulatory-Perceptual (A-P) and Conceptual-Intentional (C-I) systems respectively. The subsystem which computes $PF$ after Spell-Out is the place where the rules of the metrical theory apply to construct a complete representation of the accentual structure of the clause (Halle and Vergnaud, 1987). The analysis of the position of the main stress in
Cinque (1993) and of the focus accent in Zubizarreta (1994) are proposals to describe some syntactically-based features of this representation. At the next step, these phonological instructions are interpreted by the Articulatory-Perceptual system. The way the A-P system performs this task depends on the prosodic pattern. This is to say that the notion of prosodic pattern we introduced in this article is a notion associated with performance.

In the present paper we claim that, even if the Phonological Form depends on syntax, prosody is likely to be acquired first. Actually, the restrictions syntax imposes on the prosody give the learning child a set of hints about the grammar to be identified. This point of view is extensively developed in Morgan (1986). In his works, he argues for the following two hypotheses. First, his Bracketed Input Hypothesis says that acquisition relies on bracketed input. Second he claims that prosody plays a crucial role as a cue to bracketing. Our study of the change between CIP and EP provides evidence for these two claims.

The idea that a phonological change may drive a syntactical jump is not new. As far as we know it appears for the first time in the framework of generative grammar in the study of the change from Old to Modern French proposed in Adams (1987). Kroch (1989), shows that this point of view is supported by the quantitative analysis performed in Fontaine (1985) and is coherent with his Constant Rate Hypothesis. Kroch’s assumption that a prosodic variation may be responsible for the variation of the relative frequencies of topicalized and left-dislocated structures up to the point at which the sample is ambiguous enough to make possible a grammatical change, is equivalent to our first critical point. Moreover, even if Kroch does not explicitly consider the possibility for prosody to enter in the identification algorithm, it is natural to conjecture that the change which took place in French was driven by a change in the ratio \( \frac{p(r||p(r)}{p(r||p(r)} \) in the opposite direction of the one which took place in European Portuguese.

It is interesting, at this point, to discuss the relation between our model and Kroch’s S-shaped description of language change. In Statistical Mechanics, this type of S-shaped curves typically describes the way a metastable state relaxes to equilibrium. The initial part of the S in which the tangent is close to the horizontal corresponds to fluctuations in the magnetization due to purely random effects. When the fluctuation succeeds creating a critical droplet, the system abruptly falls in the domain of attraction of its stable state. This is described by the steep part of the curve. This description fits well with our model. Bypassing the second critical point corresponds precisely to the constitution of the critical droplet. It is important to emphasize that the S-shaped description is not particularly related to population biology models which were themselves built up by analogy with statistical mechanics models. Therefore, this description does not imply any a priori idea that grammars compete as in a darwinian picture.

In the present paper, we present a model of language acquisition driven by prosody using a pedestrian version of the Thermodynamical Formalism. This model allows us to capture the crucial interaction between competence and performance in the process of

---

28 We thank an anonymous referee for pointing out the interesting work of James Morgan to us.
29 We refer the reader to Ruelle (1978) for a general presentation of the Thermodynamical Formalism. A general discussion of Maximum Likelihood and Minimal Entropy prosody
acquisition, and understand how performance can affect grammar selection. The form of the grammar assumed by Chomsky in the Minimalist Program offers a straightforward framework for the formulation of this interaction since the levels of representation of the grammar interface with the performance systems. But while grammar is deterministic, performance is probabilistic in nature.

Acknowledgements. We thank Maria Bernadete Abaurre, Luís Carlos Cagliari, Marzio Cassandro, Pierre Collet, Artur Lopes, Gladys Massini, Jean-Roger Vergnaud and two anonymous reviewers for helpful discussions. Special thanks to Maria Aparecida Torres Moraes who provided us with unpublished data of her Ph. D. dissertation.

References.


Benincà, P. 1994. La variazione sintattica (Studi di dialettologia romanza), Bologna: Il Mulino.


Cassandro, M., A. Galves and C. Galves, in progress. Convergence time and language change. Comments on the GW model.


Chomsky, N., 1994. Bare Phrase-Structure, manuscript, MIT.


Zubizarreta, M.L., 1994. Word order, prosody and focus, manuscript, USC.

Portuguese literary works cited

(Camilo) Castelo Branco, Camilo, 1922. *Cartas de Camilo Castelo Branco a Thomaz Ribeiro* (1873-1890), Lisboa: Portugalia editora.


(Judeu) Da Silva, Antonio José 1957. *Esopaida ou a vida de Esopo*. In *Obras Com-


De Gusmão, Alexandre, s/d. Cartas (1734-1752), Lisboa: Imprensa Nacional, Casa da Moeda.


Garrett, João Batista de Almeida, s/d. Falar verdade a mentir (1845), Obras completas de Almeida Garrett, Lisboa: Empreza de história de Portugal, Sociedade Editora, Livraria Moderna.


