

The Sotaq optimality based computer program and secondary stress in two varieties of Portuguese ^{*©}

Maria Bernadete Abaurre[∞]

Charlotte Galves[∞]

Arnaldo Mandel[€]

Filomena Sandalo ^{**}

Abstract

Typical postlexical interface phenomena, like secondary (rhythmic) stress, can be successfully modeled by OT analyses, which predict optimally stressed outputs from a set of possible inputs and a hierarchically ranked set of constraints. This paper presents an OT analysis for European and Brazilian Portuguese secondary stressing. Based on this analysis, a computer program, sotaq, has been developed, allowing for automatic testing, against large corpora, of proposed constraint hierarchies for both varieties of Portuguese. Test results are presented, showing suitable hierarchies generating secondary stresses for both varieties of Portuguese.

Keywords: Optimality theory; secondary stress; European and Brazilian Portuguese; sotaq program; automated analysis; shortest path.

1. Introduction

* This research was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil) under grants 98/01831-1 and 98/3382-0 (thematic project "Rhythmic Patterns, Parameter Setting, and Language Change"), and CNPq (Conselho Nacional de Pesquisa) under grants 465928/2000-5 (TIPAL project, "Probabilistic Tools for Pattern Identification Applied to Linguistics"), 350103/92-5, and 200611/88-7.

♥ We thank Pierre Collet, João Costa, Sónia Frota, Antonio Galves and Marina Vigário for valuable suggestions and discussion of earlier versions of this paper. The annotation of secondary stresses in Brazilian Portuguese was made by Flaviane Romani and Luciani Tenani. The annotation of the European Portuguese corpus was made by João Costa, Sónia Frota and Marina Vigário. All remaining errors are of course our own. The names of the authors are arranged in alphabetical order. A first attempt to approach rhythmic differences between EP and BP in terms of constraint ranking was proposed in Abaurre & Galves (1998) and in Sandalo, Abaurre & Galves (1999).

[∞] Linguistics Department, Universidade Estadual de Campinas & CNPq.

[€] Computer Science Department, Universidade de São Paulo.

^{**} Linguistics Department, Universidade Estadual de Campinas. Corresponding author.

Present address: IEL/UNICAMP, CP 6045, Campinas-SP, Brazil 13083-970. *E-mail address* sandalo@iel.unicamp.br.

This paper presents an analysis of secondary stress in Brazilian and European Portuguese (BP and EP) based on Optimality Theory (OT). The model used associates to each linguistic string structural descriptions that are collections of decompositions into chunks of consecutive syllables, one of which is singled out as stressed. So, the input for the OT generator (*Gen*) consists of sentences and the output is a collection of feet. The secondary stresses are inferred by the OT evaluator (*Eval*) from those appearing in the segmentations that are evaluated as optimal according to a set of ranked constraints.

In current OT literature, constraint rankings have been tested manually on small data sets, with a small set of outputs. We have developed a computer program, *sotaq*, that tests proposed stress systems (formulated in terms of constraint hierarchies) for both varieties of Portuguese against observed actual stresses and in large corpora, thus allowing for automatic testing of Optimality Theory predictions for secondary stress. Such analysis is presented here, along with the corresponding constraint hierarchies for BP and EP.

2. Primary stress

Although BP and EP place primary stress exactly at the same position, secondary stress positioning is remarkably different, as can be noticed below. The examples present some possible instances of secondary stress (rhythmic stress) placement in both European and Brazilian Portuguese according to native speakers of each of the varieties. The syllables bearing primary stress are in bold and those bearing secondary stress are underlined:

EP: (1) a. A **autoridade** do **governador** diminuiu ~

- b. A autoridade do governador diminuiu
- BP: (2) a. A autoridade do governador diminuiu ~
 b. A autoridade do governador diminuiu
- EP: (3) a. A modernização foi satisfatória ~
 b. A modernização foi satisfatória
- BP: (4) a. A modernização foi satisfatória ~
 b. A modernização foi satisfatória.
- EP: (5) a. A catalogadora compreendeu o trabalho da pesquisadora ~
 b. A catalogadora compreendeu o trabalho da pesquisadora
- BP: (6) a. A catalogadora compreendeu o trabalho da pesquisadora ~
 b. A catalogadora compreendeu o trabalho da pesquisadora

The facts of primary and secondary stress in Portuguese favor Van der Hulst's (1997) position, according to which primary and secondary stresses are not derived by the same algorithm. Van der Hulst notes that, in the majority of languages, the assignment of primary stress does not depend on prior exhaustive footing. Indeed, the assignment of primary and secondary stresses in Brazilian and European Portuguese is clearly independent.

In this paper we assume that primary stress in Portuguese is part of the language's lexical information. That is, it is not assigned by the computational system of the language. Our assumption is based on the fact that, although it is well-known that Portuguese main stress falls in one of the last three syllables, none of the current analyses of Portuguese is able to successfully predict which of the three last syllables will be stressed without an extraordinary use of lexical extrametricality, as shown below.

Since many Portuguese words bear primary stress on the last syllable if it is heavy, many researchers have postulated that primary stress is assigned by constructing non-iterative moraic trochees from right to left. This is the analysis assumed, for instance, by Bisol (1992), Mateus (1975, 1983), Massini-Cagliari (1995), among many others. However, something must be said about the great number of nouns ended in light syllables that bear a stress on the last syllable (e.g. *sofá*) and about the great number of words with antepenultimate stress (e.g. *pérola*). There are also many words with penultimate stress even when the last syllable is heavy (e.g. *cadáver*). According to this analysis, most of the exceptions are dealt with via lexical extrametricality.

Given the high number of words that remain unaccounted for by an analysis that postulates moraic trochees for Portuguese, Lee (1994) revisits Camara (1953) and postulates that /e/, /a/ e /o/ in final position of nouns are thematic vowels and are outside the stress domain. According to Lee, Portuguese stress domain is the root, not the stem, and primary stressing relies on a non-iterative iambic pattern. According to this analysis, words like *mesa* bear stress on the penultimate syllable because their last vowel is a thematic vowel, that is, a suffix, and it is, therefore, outside the stress domain. And words like *sofá* bear stress on the last syllable because they do not have a thematic vowel. Although this analysis has the advantage of decreasing the number of exceptions, it is circular because we only know that a vowel is thematic (i.e. a suffix) once we know whether it is stressed. In addition to its circularity, this analysis still has many exceptions since the words with an antepenultimate stress pattern and the words ending by a heavy syllable bearing a penultimate stress pattern remain unaccounted for.

In conclusion, both types of analysis require an extraordinary amount of lexical extrametricality to solve the great number of exceptions, which suggests that it is more

economical to postulate that primary stress is phonemic. This kind of conclusion is already widely assumed for Spanish, whose main stress phenomena are quite similar to Portuguese. According to Hayes (1995:96), "main stress in Spanish is phonemic, though it can be predicted to a fair extent by complex lexical rules, whose character continues to be debated".

3. Secondary stress

Our analysis of secondary stress is based on a corpus of 20 sentences which were read three times by three native speakers of Portuguese from Lisbon, Portugal, and by two native speakers of Portuguese from São Paulo, Brazil. The data have been transcribed on the basis of auditive perception, but spectrograms were used as support for the phonetic transcription¹.

Our analysis holds for a normal rate of speech in sentences that convey new information, as in headline news. Slow, deliberate speech can lead to stress patterns that will be disregarded here. For instance, it is well known that a different stress pattern may result from what intuitively feels like special emphasis on a particular element.

The data suggested two basic distinctions:

- In BP, secondary stress follows a binary pattern, while no similar restriction holds for EP;

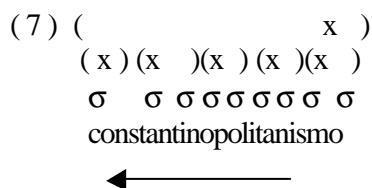
¹ The spectrograms were made by means of the Multi Speech program for acoustic analysis.

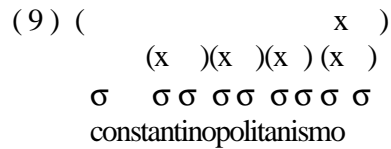
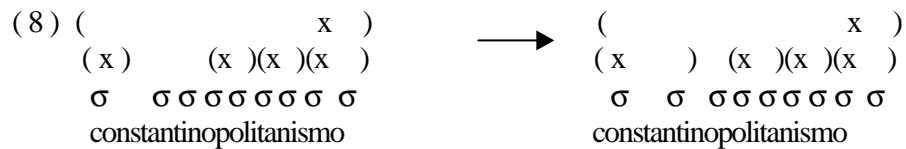
- EP allows functional words to be stressed, while BP does not.

We elaborate on this as follows.

Brazilian Portuguese secondary stress follows a rarely violated binary (two-syllable) pattern. The exceptions to the binary system are mostly cases of the so-called initial dactyl (Prince 1983). That is, there is an initial ternary alternation (the initial dactyl) when the stress domain has an odd number of syllables. The initial dactyl is not obligatory, however. For instance, a word like *satisfatória* can be stressed as *satisfatória*, an example of the initial dactyl, or as *satisfatória*.

It is well known that Spanish presents exactly the same phenomenon (Harris 1983, 1989, Roca 1986). Harris (1989), within Metrical Theory, has suggested an analysis for Spanish which states that the two variants represent alternative outcomes to the resolution of a stress clash. On Harris's analysis, secondary stress in Spanish is applied by building trochees from right to left on the syllables preceding the syllable bearing main stress. If we allow degenerate feet at an intermediate stage of the derivation, the sort of clash shown in 7 will result. Initial dactyls can then be derived by applying a rule of rightward destressing and reparsing, whose effects are shown in 8, where one syllable in the middle of the word (*ti*) is left unparsed. The other option is to resolve the clash with leftward destressing, as shown in 9.





Hayes (1995) points out that "the crucial point of Harris's analysis is that it relies on a temporary degenerate foot, set up in the middle of the derivation (7), that either is expanded into a proper foot by destressing and reparsing, or is itself deleted." In neither case the degenerate foot surfaces and Hayes maintains that it shows that the crucial point of the Spanish phonology is the presence of a constraint that bans degenerate feet.

One could argue that the same analysis could be employed for BP. Our data, however, shows that this analysis faces empirical problems, as discussed below.

An acoustic analysis of the BP facts shows that many words containing an odd number of syllables have undergone vowel deletion, which resulted in a perfect binary system. In other words, the syllable that Harris supposes to be left unparsed is actually not realized. Thus, the word *satisfatória* was actually realized as *satsfatória*, where the vowel /i/ has been deleted, resulting in a perfect binary structure ((satsfa) Σ (tória) Σ). One could argue that the one strategy employed for Brazilian Portuguese to avoid degenerate feet is vowel deletion instead of simply reparsing. Thus, an analysis along the lines of Harris's proposal could be offered, provided that a rule of /i/ deletion is added. The phenomenon of vowel deletion in Brazilian Portuguese, however, shows that the facts are more complex than a metrical analysis can predict. The words containing an odd number of syllables are

the target for vowel deletion, which suggests that we are indeed looking at a language that prefers to avoid degenerate feet, as claimed by Hayes. The realizations in 10 and 11, however, are problematic for Metrical Theory because, if secondary stress results from an alternation of stressed and non-stressed syllables from right to left on the syllables preceding the syllable bearing main stress, there would be no reason for vowel deletion because there are four syllables preceding the syllable with main stress in *investigador* and in *modernização*, and therefore a perfectly binary alternation would result. The prosodically-induced vowel deletion of 10 and 11 only makes sense if we assume that there is a constraint that forces binary feet (i.e. (in vest)_Σ (ga **dor**)_Σ) and (mo dern)_Σ (za **ção**)_Σ), and there is no need to introduce directionality (right to left counting) in order to obtain binarity via perfect alternation between strong and weak syllables, as predicted by a Metrical Theory analysis.

(10) O in ves ti ga dor já lhe de vol veu o di nhei ro.

[win vest ga **dor** já lhe de vow **vew:** di **nhei** ro]

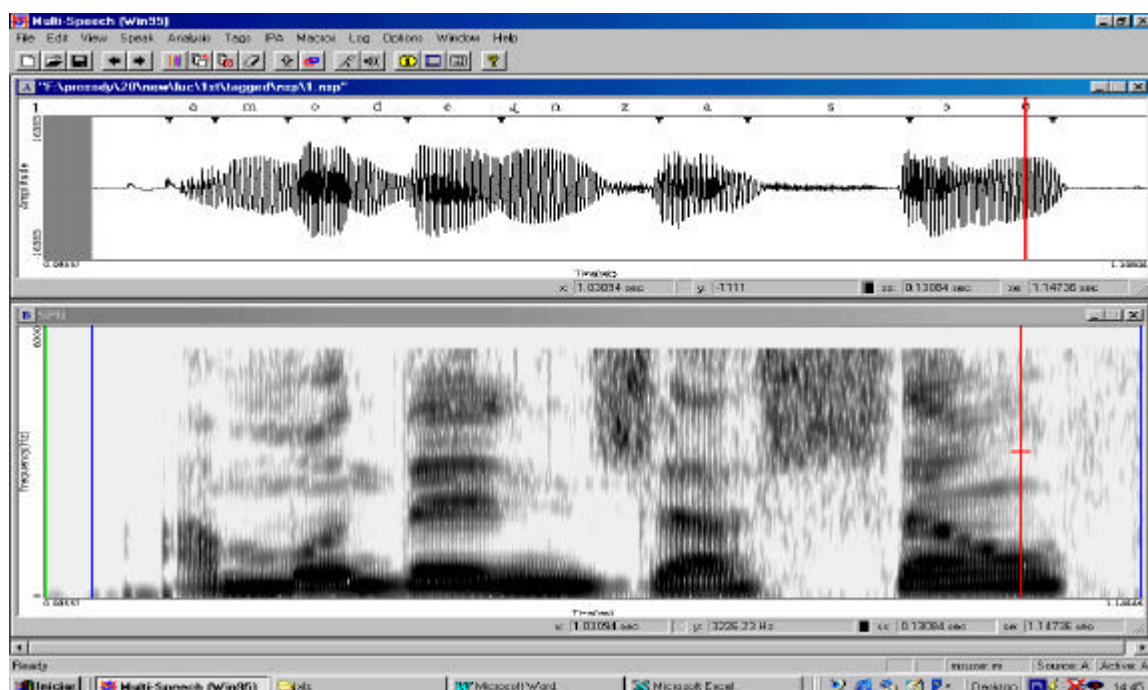
(11) A modernização foi satisfatória

[a mo dern za **ção** foi sats fa **tó** ria]

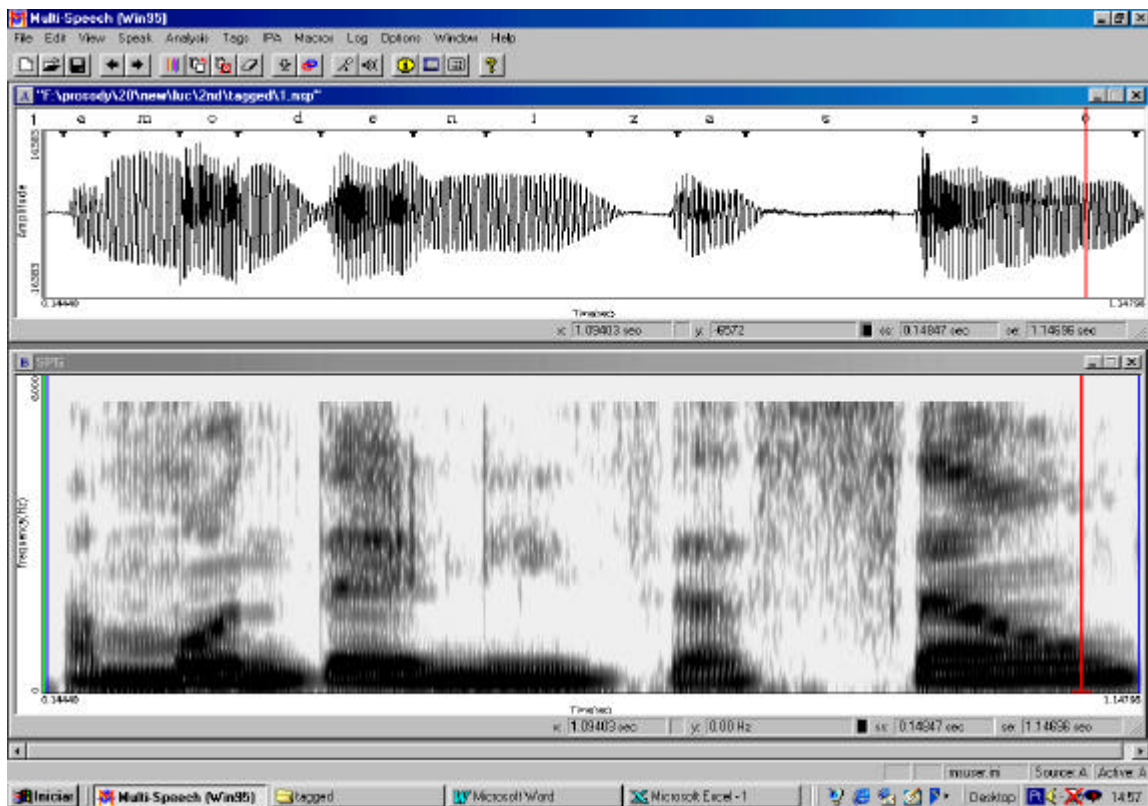
We will propose in section 3 that the facts of Portuguese result from a conflict of forces instead of from a computation of alternating strong and weak syllables like it has been widely assumed for Spanish and also for Brazilian Portuguese within Metrical Theory (see for instance Collischonn 1993 for Brazilian Portuguese). In this system we will derive the facts of initial dactyl without postulating degenerate feet that never surface. Such degenerate feet represent cases of absolute neutralization and it is widely accepted that

absolute neutralization must be avoided given the problems that it may bring for language acquisition. Since our OT analysis makes it possible to generate cases of initial dactyl where there are no cases of vowel deletion, it may be the case that our analysis can be extended also to Spanish avoiding absolute neutralization also for that language.

A process of vowel deletion that forces a binary system has been noticed before for primary stress (Bisól 2000, among others). For instance, it is well known that words like *pérola* ‘pearl’ are often realized as *perla*. This paper represents the first time that a similar phenomenon has been noticed for secondary stressing. Abaurre (1979) discusses several cases of vowel deletion in BP, but the phenomenon is not associated with foot binarity. Below are the acoustic configurations of the word *modernização*, where the first spectrogram attests the mentioned vowel deletion and the second spectrogram shows the



same word with no vowel deletion.



European Portuguese differs from Brazilian Portuguese in that it is not a binary system. In European Portuguese the beginning of a sentence tends to be prominent, as noticed already by Frota (1998) and Vigário (1998). This fact can be noticed in the example below:

(12) O **investiga** **dor** **já** me ofere**ceu** **dinheiro** ~

O investiga **dor** **já** me ofere**ceu** **dinheiro**.

But we find in our corpus other prominences at the beginning of smaller domains (cf. A **cataloga** **dora** compre**deu** o **trabalho** da **pesquisadora**). We leave for further research the exact characterization of this domain. Here, we refer to such domain as phonological

phrase.² The important fact is that EP shows unbounded secondary footing. D'Andrade & Laks (1991) have claimed that secondary stresses are assigned via binary feet construction in EP, and Carvalho (1988/1989) claims that secondary stress is assigned via ternary feet. The transcription of our data by three native speakers of EP does not indicate either binary or ternary alternations.

Another point where EP and BP differ concerning secondary stressing is that functional words can bear secondary stress in EP (A *catalogadora* ~ A *catalogadora*). That is, EP accepts the placement of a secondary stress on either the functional word that starts a phonological phrase or on the first syllable of the first lexical word of a phonological phrase. In BP, functional words never bear stress in a non-emphatic pronunciation. Finally, EP and BP differ in that only EP has the option of not assigning any secondary prominences in a word (cf. O *investiga***do**r já lhe devolve**u** o *di***n**heiro).

The variation on secondary stress placement in both EP and BP is problematic for a Metrical Theory analysis because, in a derivational analysis, we would have to postulate that one form is default and derive the other form via re-arrangement rules. Since EP accepts a range of variation that includes even the possibility of not assigning any secondary stresses, the re-arrangement rules for EP could be so complex as to make a derivational analysis unwieldy.

To sum up, an analysis in OT terms has the advantages of : (i) generating all the facts of both Brazilian and European Portuguese without postulating any cases of absolute neutralization; (ii) not forcing the usage of the notion of directionality, thus implying a

² According to Vigário (personal communication), the relevant domain for EP is the phonological word, rather than the phonological phrase. This will be taken into account in future developments of this work.

simplification of the phonological theory; and (iii) being able to generate variant forms in parallel.

3. An Optimality Analysis

We now describe our OT model in detail.

The inputs will be sentences in a language (in our case, BP or EP). The structures assigned by *Gen* to each input are decompositions into *segments*. Those objects can be of two types:

- 1) A *regular segment* - a sequence of consecutive syllables, with one of them marked. This marked syllable is called the *core* of the segment. Note that this is a formal construct: the tagging has no a priori relation to any stresses, primary or secondary.
- 2) A *pseudo segment* - a single syllable, with no stress.

Furthermore, in a segmentation yielded by *Gen*, each syllable is contained in exactly one segment.

Recall that, in constructing an OT tableau, one draws one line for each possible output, that is, for each segmentation, in our case. As we will see later, tableaux are totally impractical for this model, since even moderately sized inputs have an extremely large number of possible outputs. Even a computer would not be able to list all those outputs, so a true mathematical optimization approach has to be taken to find the true optimal solutions without exhaustively searching all possibilities.

Each segmentation yields a stressing of the given sentence simply by stressing the core of the regular segments. Under this correspondence, regular segments can be interpreted as metric feet of the resulting stressed sentence.

This model entails a specific locality restriction on the type of constraints we are willing to consider: each constraint ought to be checkable by considering each segment individually, *or* by checking each pair of adjacent segments. It turns out that most constraints already used in other OT work can be expressed this way, so we are not handicapping ourselves too much.

One important aspect of our model is that we have not restricted ourselves to a strict ranking of the constraints, but have completely accepted the possibility of a stratified dominance hierarchy. The reason for that is large amount of free variation observed in our data, and the virtual impossibility of accounting for it with strict hierarchies. We note that the way we use stratified hierarchies differs from the proposal of Reynolds (1994), and is not exactly contemplated in section 9.4 of Kager (1999); the counting is reminiscent of the Tesar/Smolensky Learning Algorithm. Our data gives the impression that we can better account for free variation than by other methods, although further research is needed to clarify whether this is the case.

We describe each constraint that follows in two forms: An intensional form (in italics), giving an idea, and a formal form, telling when a violation mark must be assigned.

The constraints found to be relevant to this analysis so far are:

DEPST: *Deletion of lexical stresses is not allowed.* Violated by a segment containing a

lexically stressed syllable not tagged as the core.

RIGHTMOST: *Lexical stresses occur at the last foot of each lexical word.* Violated by a

segment not containing the last syllable of a word, provided the segment's core has a lexical stress.

INTLEX: *A lexical word must be a prosodic word.* Violated by a segment that contains syllables of more than one word.

ALIGN (FT, L, PHP L): *Every foot has its left boundary at the left edge of a phonological phrase.* Violated by a regular segment whose left boundary is not the left edge of a phonological phrase.³

FOOTBIN/BINGRAD: *Feet must be binary.* FOOTBIN is violated by a regular segment that does not have exactly two syllables. BINGRAD is a gradient form of the same restriction: long feet count one violation for each syllable exceeding the initial two.

PARSE: *All syllables must be parsed into feet.* Violated by each pseudo-segment which is not a functional word.

TROCHEE: *All feet must be left-headed.* Violated by a segment whose core is not its initial syllable.

NOCLASH: *No stressed syllables can be adjacent.* Violated by a pair of successive segments, the core of the first in the last syllable, the core of the second in its initial syllable.

CLASHINT: *No stressed syllables within a lexical word can be adjacent.* Like NOCLASH, but with the two cores in the same word.

CLASHEXT: *No stressed syllables in successive words can be adjacent.* Like NOCLASH, but with the second segment starting at the beginning of a lexical word.

NOLAPSE: *No adjacent unstressed syllables inside a word-medial foot.* Violated by a

segment occurring not at the beginning or end of a lexical word containing two adjacent non-core syllables.

One should get an intuitive picture of the dynamic of conflict among those constraints, in order to appreciate their significance.

Two of them, RIGHTMOST and CLASHINT, are undominated in both EP and BP. This is evident in the analyzed corpus, and we conjecture that this is a valid generalization for these languages. RIGHTMOST reflects properties of lexical words.⁴ CLASHINT comes from the rhythmic need for alternation of prominent and non-prominent syllables in these languages for the style of speech we are considering.

Next comes DEPST, which for the experiments we have done has surfaced as undominated. There are, however, circumstances of primary stress retraction in PB, as discussed by Sandalo & Truckenbrodt (2001). In the OT system, this fact can be captured by other constraints that are not listed here and that dominate DEPST. In future developments of this work, these constraints will be explored.

In opposition to FOOTBIN/BINGRAD, NOLAPSE, and INTLEX (that tend to require small segments), there are Parse and ALIGN (FT, L, PHP L).⁵ The latter constraint forces the inclusion of a functional word in the stress domain as well as forces unbounded feet since it requires all feet to be aligned with the phonological phrase. The different

³ In the actual version of sotaq, phonological phrases are delimited manually, but see Sandalo & Truckenbrodt (2001) for a proposal that could be implemented in sotaq in order to generate phonological phrases in BP automatically. See also footnote 2 for EP.

⁴ Indeed, this constraint is so obviously true that we only were made aware of the need to make it explicit when Sotaq started producing unacceptable segmentations that clearly violated it. It is well known that all known natural languages have primary stress on an edge window, and Rightmost just places the window for Portuguese.

⁵ Intlex is given as *LXWD = PRWD* in Kager (1999).

hierarchization of these constraints will be crucial to derive the main differences between BP and EP. In BP foot binarity is crucial and in EP feet are unbounded.

Another type of opposition is that between TROCHEE and CLASH EXT. While TROCHEE forces a specific position for the stressed syllable in a foot, CLASH EXT disallows some positions due to the interaction of adjacent feet.

Note that, following the assumptions of OT, we use the same constraints with different rankings to derive BP and EP stress patterns. There is, however, a noticeable partial exception, namely FOOTBIN/BINGRAD.

All the constraints, except for one, have categorical violations. FOOTBIN/BINGRAD is a manifestation of a same constraint with different ways to compute violation. While violations of FOOTBIN are computed as categorical, violations of BINGRAD are gradient. Recall that a long foot will compute as a single violation of FOOTBIN, whereas for BINGRAD the number of violations increases with the length of the foot. The strong preference for binary feet in BP has been attested in many works (Bisol 1992, Collischonn 1993, Lee 1994, Massini-Cagliari 1995). Moreover, our handling of the data showed that FootBin is too weak a constraint for generating the correct facts of BP, while BinGrad is too strong for EP, even if very lowly ranked.⁶

Recall that BP shows a phenomenon of vowel deletion induced by rhythm. It is well known that EP also undergoes vowel deletion (Mateus 1975, 1983). We could not employ constraints to handle the BP/EP facts because the phenomenon of vowel deletion has not been fully understood, for both Portuguese varieties, until the present point of this research. In other words, we had to avoid handling vowel deletion automatically at this point of our

project because we do not have a complete description of the facts yet. It is, however, crucial to implement the BP rhythmic vowel deletion in order to generate the correct facts relative to secondary stress in this variety of Portuguese. Therefore, we informed sotaq manually what we know about vowel deletion (via spectrogram analysis of our corpus). We marked with a + the BP vowels that can be deleted. The program does not count as a violation of BINGRAD any foot containing three syllables one of which contains a vowel marked by +.

Other processes of resyllabification, like those resulting from the application of vocalic sandhi rules, or internal diphthongization are also important, and our tests and data clearly shows so. At this moment we have not dealt with those phenomena yet, but they will be duly considered in the future.

⁶ The implications of this fact for language acquisition remain to be investigated. It may be the case that specific the way (gradient or categorical) to compute violations of a constraint is not innate, but acquired via exposition to the input.

4. Sotaq

In most of the current OT literature, a given analysis is tested manually via manipulation of a very restricted amount of data, usually consisting of words or very short phrases. We have developed a computer program, named sotaq, for automatic testing of various different constraint hierarchies on a robust amount of data, thus providing more substantive evidence for our analysis. An earliest version of such a program was conceived and implemented by Pierre Collet and Antonio Galves; the current version is a new implementation, building on their initial ideas.

Here we present an abridged description of sotaq, explaining some of its underlying algorithms. A more detailed explanation will be presented elsewhere.

Roughly speaking, sotaq is fed a constraint hierarchy and it processes sentences assigning secondary stresses according to the corresponding OT model. The constraints explained earlier are all implemented. As can be seen from their definition, their computation requires information about syllabification, lexical stresses, lexical words and phonological phrases. While much of this information could be computed automatically from the sentences, at this point they are given in the input. This complicates the input slightly, but makes sotaq a leaner program, concentrated on its main task.

The input for sotaq is a collection of sentences, each one being a collection of *tagged syllables*. Each of these is a phonological syllable in an actual Portuguese utterance, preceded by a numerical tag that encodes some properties of that syllable. Some such properties are: whether it starts a word, whether it has a primary stress, whether a vowel can be erased in speech. Here is an example:

6 0 10 in 0 ves 16 ti 0 ga 1 dor 7 já 2 lhe 10 de 0 vol 1
veu 6 o 2 di 1 nhei 0 ro.

The tag for a syllable is a sum of values as such:

O: 2 (starts a word) + 4 (starts a phonological phrase)

in: 2 (starts a word) + 8 (secondary stress⁷)

já: 1 (primary stress) + 2 (starts a word) + 4 (starts a phonological phrase)

ti: 16 (vowel may be erased).

The constraints may refer to the tags; actually, *sotaq* processes only the tags in its search for the optimal stresses; the textual syllables are used only to produce human-readable output⁸.

When the program is called, the name of the file containing sentences tagged as above is specified, together with a constraint hierarchy. Here are two examples of *sotaq*'s output; we use the same input sentence, with two different rankings.

Example 1: The ranking used was:

```
DEPST : RIGHTMOST : CLASHINT >> INTLEX >> BINGRAD : PARSE : NO LAPSE >> CLASHEXT  
>>TROCHEE >> ALIGN
```

This is the one that best fits BP, so far. The output was:

```
I: A in·te·li·gên·cia da ca·ta·lo·ga·do·ra foi  
de·ter·mi·nan·te  
  [ |  ~ ~    ^^^    [ | ~ ~    ~ ~    ^^    [ ^^^ |    ~ ~ ~    ^^^  
+  
O: |a |in|TEli|GÊNcia |da |CAta|LOga|DOra |FOI |deTER|miNANte  
[0]
```

⁷ This secondary stresses were also obtained through auditory perception, and the transcriptions can be used to check *sotaq*'s result. This way, one may test the OT model's predictiveness.

⁸ Currently there are no user-friendly facilities for preparing input for *sotaq*. A better user-interface is already being designed, and will be implemented RSN.

O: |a |INteli|GÊNcia |da |CAta|LOga|DOra |FOI |deTER|miNANte
[2]

Total cost: 2027.

19 syllables.

2 optimal segmentations.

+147667381 possible segmentations (tableau lines)

Example 2: The ranking used was:

DEPST : RIGHTMOST : CLASHINT >> TROCHEE >> ALIGN : INTLEX >> PARSE >> CLASHEXT
>> FOOTBIN : NOLAPSE

This is the one that best fits EP, so far. The output was:

I: A in·te·li·gên·cia da ca·ta·lo·ga·do·ra foi
de·ter·mi·nan·te
[|~~ ^^^ [|~~ ^^ [^^^| ^^^
O: |a |INteli|GÊNcia |da |CATaloga|DOra |FOI determi|NANte
[0]
O: |A inteli|GÊNcia |da |CATaloga|DOra |FOI determi|NANte
[2]
O: |a |INteli|GÊNcia |DA cataloga|DOra |FOI determi|NANte
[2]
O: |A inteli|GÊNcia |DA cataloga|DOra |FOI determi|NANte [4]

Total cost: 6003.

19 syllables.

4 optimal segmentations.

+147667381 possible segmentations (tableau lines)

In each case, the line labeled I and the following one describe the input. The first line is the textual sentence, the second describes tags. A [marks the beginning of a phonological phrase, a | marks the beginning of a lexical word; a syllable is underlined with carets (^^^) if it bears a primary stress, and it is underlined with tildes (~~~) if secondary stress was auditorily perceived. Further, a + under a vowel means that it may be deleted in speech.

| | | | | | | | | |
|------------------------|---|---|---|---|---|---|---|---|
| lOga | | | | | | | | * |
| DORa | | | | | | | | * |
| FOI | | | | * | | | | |
| deTER | | | | | | | * | * |
| miNAN | | | | | | | * | * |
| te | | | | | | | | |
| viola tion count | 0 | 0 | 1 | 0 | 2 | 6 | | |

Note that the last segment, miNANte, even though it contains three input syllables, does not violate BINGRAD. That is because the vowel in the last syllable, te, has been tagged as erasable, so sotaq predicts its actual deletion in speech.

The two solutions differ in how the three syllables in-te-li are parsed into segments. In the first solution, we see in|TEli. Here, in violates PARSE, and TEli violates align. In the second solution, we see INteli. This one violates BINGRAD and ALIGN. Since PARSE and BINGRAD are equally ranked, those two segmentations have the same violation count. So, at the end, the violation count, following the columns in the table, is 0,0,2,0,2,7. This sequence is encoded in the total cost reported by sotaq, 2027 in this case.

It is worth noticing how the attested variety can only be achieved here by virtue of the equal rank of PARSE and BINGRAD. If one insisted in a strict ranking, only one of the two solutions would have been optimal, and the other would be lost.

Sotaq is entirely written in perl, so it is supposed to run on any system where perl is installed. We have only used it in Unix-like systems, however. It is normally called as

```
sotaq --rank xxxx file
```

where `xxxx` is a description of a constraint hierarchy, and `file` is the name of an input file containing tagged phrases. The `xxxx` for example 1 above was `depst:rightmost:clashint%intlex%parse:bingrad:nolapse%clashext%trochee%align`. Other command line options exist for more control on the output, experimentation with some nonstandard constraints, and also for experimentation with somewhat fuzzy hierarchies (a simple mathematical idea, with no current linguistic support).

A word on sotaq's innards. The possible segments are treated as nodes in an acyclic directed graph, so that a segmentation becomes a directed path between the single source and the single sink. The constraint hierarchy yields positive real costs assigned to nodes and edges, in such a way that finding the OT-preferred segmentation becomes the problem of finding shortest source-sink paths. That is done by a variation of Dijkstra's Algorithm, with few implementation optimizations, as we do not expect to run sotaq on overly long utterances.

5. Testing a corpus

There follows the list of twenty phrases that was used as our data for the proposed analysis, tested by sotaq for both EP and BP. In addition to the 20 sentences, we also include a fragment of a text tested by sotaq for both EP and BP. Note that no deletable

vowel was marked in the text fragment, yet. The stresses assigned, thus, correspond to sotaq's prediction for secondary stress in a situation where vowel deletion is avoided. Note, also, that in the first batch of tests we have not required sotaq to separate printed feet by bars.

We present now three tests of sotaq in different corpora. In all cases the rankings used were those in the examples in the last section, using that of Example 1 whenever we wanted to test for BP, and that of Example 2 whenever EP was the case.

Test 1 is based on a corpus of 20 phrases. Each was read by a speaker of each variety of Portuguese, and the observed secondary stresses were annotated. In this way, one can actually get an impression of the adequacy of the model to reality. However, no hasty conclusions should be arrived at from so small a sample.

Test 2 is based on a text fragment, with long sentences, which also was read by one native BP speaker, with the input for sotaq correspondingly annotated. We have no corresponding EP reading, so, in the unannotated run using the EP hierarchy, the brackets at the right simply tell how many secondary accents were predicted by sotaq.

Test 3 tries to work out sotaq limits on hard sentences, in the sense that they stress the algorithm and its assumptions. The sentences are somewhat nonsensical, but morphologically and syntactically Portuguese, so they can be read quite naturally; we have not, however, annotated them based on actual readings. Testing sotaq on this type of data made us aware of some subtle interplay of the constraints, and was crucial in finally figuring out the constraints and rankings that best modeled our corpora.

The run results were slightly reformatted, since the long lines sotaq produces would be hard to read in print.

Test 1

```
[frases] > ../sotaq --count --rank
depst:rightmost:clashint%intlex%bingrad:parse:nolapse%clashext%trochee%align lu
```

Active constraints, absolute weights, relative weights:

```
align:          1
bin grad:       1000
dep st:         100000
lex integr:     10000
nolapse:        1000
parse:          1000
rightmost:     100000
trochee:        10
ext noclash:   100
int noclash:   100000
```

```
# Frase 1
# Lu2-11.wav
#
```

```
I:  A  mo · der · ni · za · ção  foi  sa · tis · fa · tó
    · ria
```

```
      [  |  ~ ~      +      ^ ^ ^ [ ^ ^ ^ |  ~ ~      +      ^ ^
O:  |a |MODerni|zaÇÃO |FOI |SATisfa|TÓria [0]
```

Total cost: 1214.

12 syllables.

1 optimal segmentation.

+51092 possible segmentations (tableau lines).

```
#
# Frase 2
# Lu2-11.wav
#
```

```
I:  A  au · to · ri · da · de  do  go · ver · na · dor
    di · mi ·
```

```
      [  |  ~ ~      ^ ^      + [  |  ~ ~      ^ ^ ^ [
~ ~
    nu · iu
      ^ ^
```

```
O:  |a |AUto|riDAde |do |GOver|naDOR |diMI|nuIU [0]
```

Total cost: 145.

15 syllables.

1 optimal segmentation.

+2565168 possible segmentations (tableau lines).

Frase 3
Lu2-11.wav
#

I: O in · ves · ti · ga · dor já de · vol · veu · o
di ·

[| ~ ~ + ^ ^ ^ [^ ^ | ~ ~ ^ ^ ^ + |
nhei · ro
^ ^ ^ ^

O: |o |INvesti|gaDOR |JÁ |DEvol|VEUo |di|NHEIro [0]

Total cost: 2215.

14 syllables.

1 optimal segmentation.

+400542 possible segmentations (tableau lines).

Frase 4
Lu2-11.wav
#

I: O or · ga · ni · za · dor a · pre · sen · tou a
ca ·

[| ~ ~ ~ + ^ ^ ^ [~ ~ ~ ^ ^ ^ [|
~ ~
ta · lo · ga · do · ra
~ ~ ^ ^

O: |o |ORgani|zaDOR |aPRE|senTOU |a |CAta|LOga|DOra [1]

Total cost: 36.

17 syllables.

1 optimal segmentation.

+20383750 possible segmentations (tableau lines).

Frase 5
Lu2-11.wav
#

I: A fal · ta de mo · der · ni · za · ção é ca ·
tas ·

[| ^ ^ ^ [| ~ ~ + ^ ^ ^ [^ | ~ ~
tró · fi · ca
^ ^ ^ +

O: |a |FALta |de |MODerni|zaÇÃO |É |CATas|TRÓfica [0]

Total cost: 1215.

15 syllables.

1 optimal segmentation.

+1167180 possible segmentations (tableau lines).

#

Frase 6

Lu2-11.wav

#

I: O tra · ba · lho da pes · qui · sa · do · ra foi
pu

[| ^ ^ [| ~ ~ ~ + ^ ^ [^ ^ ^
| ~ ~
· bli · ca · do
^ ^

O: |o |tra|BALho |da |PESquisa|DOra |FOI |PUBli|CAdo [0]

Total cost: 2105.

15 syllables.

1 optimal segmentation.

+2861585 possible segmentations (tableau lines).

#

Frase 7

Lu2-11.wav

#

I: O pro · fe · ssor tam · bém o des · cre · veu
aos

[| ~ ~ ~ ^ ^ ^ ^ [^ ^ ^ [| + ^ ^ ^ [
|
a · lu · nos
^ ^

O: |o |PROfe|SSOR |tamBÉM |o |descreVEU |aos |a|LUNos [0]

Total cost: 2024.

14 syllables.

1 optimal segmentation.

+1074186 possible segmentations (tableau lines).

#

Frase 8

Lu2-11.wav

#

I: O go · ver · na · dor a · cei · tou a mo · der ·
ni ·

[| ~ ~ ^^^ [^^^ [| ~ ~

~+

za · ção
 ^^^

O: |o |GOVer|naDOR |aceiTou |a |MOderni|zaÇÃO [1]

Total cost: 1134.

14 syllables.

1 optimal segmentation.

+765282 possible segmentations (tableau lines).

#

Frase 9

Lu2-11.wav

#

I: A fal · ta de au · to · ri · da · de foi a ·
lar ·

[| ^^^ [| ~ ~ ^^ + [^^^ | ~

man · te

^^^ +

O: |a |FALta |de |AUto|riDAde |FOI |Alar|MANTe [2]

Total cost: 1115.

14 syllables.

1 optimal segmentation.

+898025 possible segmentations (tableau lines).

#

Frase 10

Lu2-11.wav

#

I: O in · ves · ti · ga · dor já lhe de · vol · veu
o

[| ~ ~ + ^^^ [^^ | | ~ ~ ^^^

[|
di · nhei · ro
 ^^^^

O: |o |INvesti|gaDOR |JÁ |lhe |DEvol|VEU |o |di|NHEIro [0]

Total cost: 3115.

15 syllables.

1 optimal segmentation.

+1385268 possible segmentations (tableau lines).

Frase 11
Lu2-11.wav
#

I: A ca · ta · lo · ga · do · ra com · preen · deu o
tra
[| ~ ~ ~ ~ ^ ^ [~ ~ ~ ~ ^ ^ ^ [
|
· ba · lho da pes · qui · sa · do · ra
^ ^ [| ~ + ~ ^ ^
O: |a |CAta|LOga|DOra |COMpreen|DEU |o |tra|BALho |da
|PESquisa|DOra [2]

Total cost: 2007.
20 syllables.
1 optimal segmentation.
+672812904 possible segmentations (tableau lines).

Frase 12
Lu2-11.wav
#

I: A pro · fe · sso · ra dis · cu · tiu a gra · ma
· ti ·
[| ~ ~ ~ ^ ^ ^ [+ ^ ^ ^ [| ~ ~
+
ca · li · da · de
^ ^ +
O: |a |PROfe|SSOra |discuTIU |a |GRAMati|CALi|DAde [3]

Total cost: 15.
16 syllables.
1 optimal segmentation.
+9475728 possible segmentations (tableau lines).

Frase 13
Lu2-11.wav
#

I: A in · te · li · gên · cia da ca · ta · lo · ga ·
do · ra

[| ~ ~ ^ ^ ^ [| ~ ~ ~

foi de · ter · mi · nan · te

[^ ^ ^ | ~ ~ ~ ^ ^ ^ +

O: |a |in|TEli|GÊncia |da |CAta|LOga|DOra |FOI
|deTER|miNANte [0]

O: |a |INteli|GÊncia |da |CAta|LOga|DOra |FOI |deTER|miNANte
[2]

Total cost: 2027.

19 syllables.

2 optimal segmentations.

+147667381 possible segmentations (tableau lines).

#

Frase 14

Lu2-11.wav

#

I: O in · ves · ti · ga · dor já o · fe · re · ceu
di ·

[| ~ ~ ~+ ^ ^ ^ [^ ^ | ~ ~ ~ ^ ^ ^ [

nhei · ro

^ ^ ^ ^

O: |o |INvesti|gaDOR |JÁ |oFE|reCEU |di|NHEIro [1]

Total cost: 2135.

14 syllables.

1 optimal segmentation.

+302214 possible segmentations (tableau lines).

#

Frase 15

Lu2-11.wav

#

I: A pro · cu · ra da gra · ma · ti · ca · li · da ·
de é

[| ^ [| ~ ~ ~ + ~ ~ ^ ^

[^ |

o no · sso o · bi · je · ti · vo

| ^ ^ [~ + ^ ^

O: |a |pro|CUra |da |GRAMati|CALi|DAde |É |o |NOSso
|Obije|Tivo [2]

Total cost: 2006.

21 syllables.

1 optimal segmentation.
+1071019822 possible segmentations (tableau lines).

Frase 16
Lu2-11.wav
#

I: A pes · qui · sa · do · ra per · deu au · to · ri
· da · de

[| ~~~ + ^^ [^^^ [~~
^^ +

O: |a |PESquisa|DOra |perDEU |auTO|riDAde [0]

Total cost: 33.
13 syllables.
1 optimal segmentation.
+391005 possible segmentations (tableau lines).

Frase 17
Lu2-11.wav
#

I: O pro · fe · ssor tam · bém des · cre · veu os

a ·
[| ~~~ ^^^^ [^^^ | + ^^^ [|

lu · nos
^^ +

O: |o |PROfe|SSOR |tamBÉM |descreVEU |os |aLUnos [0]

Total cost: 1034.
13 syllables.
1 optimal segmentation.
+312469 possible segmentations (tableau lines).

Frase 18
Lu2-11.wav
#

I: A au · to · ri · da · de ca · be ao go · ver ·
na · dor

[| ~~ ^^ + [^^ [| ~~~
^^^

O: |a |AUto|riDAde |CAbe |ao |GOver|naDOR [2]

Total cost: 124.
13 syllables.
1 optimal segmentation.
+299349 possible segmentations (tableau lines).

Frase 19
Lu2-11.wav
#

I: 0 in · ves · ti · ga · dor já me o · fe · re ·
ceu di
[| ~ ~ + ^ ^ ^ [^ ^ | | ~ ~
^ ^ ^ [
· nhei · ro
^ ^ ^ ^

O: |o |INvesti|gaDOR |JÁ |me |Ofe|reCEU |di|NHEIro [2]

Total cost: 2125.
15 syllables.
1 optimal segmentation.
+991233 possible segmentations (tableau lines).

Frase 20
Lu2-11.wav
#

I: A gra · ma · ti · ca · li · da · de das fra · ses
foi
[| ~ ~ + ^ ^ [| ^ ^ ^
[^ ^ ^ |
con · se · gui · da
~ ~ ~ ^ ^ ^

O: |a |GRAMati|CALi|DAde |das |FRases |FOI |CONse|GUIDa [3]

Total cost: 1106.
16 syllables.
1 optimal segmentation.
+5559320 possible segmentations (tableau lines).

[frases] > ../sotaq --count --rank
depst:rightmost:clashint%trochee%align:intlex%parse%clashext%
footbin:nolapse marina

Active constraints, absolute weights, relative weights:
align: 1000
dep st: 100000

binary foot: 1
lex integr: 1000
nolapse: 1
parse: 100
rightmost: 100000
trochee: 10000
ext noclash: 10
int noclash: 100000

Frase 1
Marina2-11.wav
#

I: A mo · der · ni · za · çã o foi sa · tis · fa · tó
· ria

[~ | ^^^ [^^^ | ^^

O: |A moderniza|ÇÃO |FOI satisfa|TÓria [0]

O: |a |MOderniza|ÇÃO |FOI satisfa|TÓria [2]

Total cost: 4013.
12 syllables.
2 optimal segmentations.
+51092 possible segmentations (tableau lines).

Frase 2
Marina2-11.wav
#

I: A au · to · ri · da · de do go · ver · na · dor
di ·

[| ~ ~ ^ ^ [| ^^^ [

+
mi · nu · iu
^^

O: |A autori|DAde |DO governa|DOR |DIminu|IU [4]

O: |a |AUtori|DAde |do |GOverna|DOR |DIminu|IU [2]

O: |A autori|DAde |do |GOverna|DOR |DIminu|IU [4]

O: |a |AUtori|DAde |DO governa|DOR |DIminu|IU [2]

Total cost: 5025.
15 syllables.
4 optimal segmentations.
+2565168 possible segmentations (tableau lines).

Frase 3

Marina2-11.wav
#

I: O in · ves · ti · ga · dor já de · vol · veu o
di ·
[| ~ ~ ^^^ [^ ^ | ^^^ [|
nhei · ro
^^^^

O: |o |INvestiga|DOR |JÁ devol|VEU |o |di|NHEIro [0]
O: |O investiga|DOR |JÁ devol|VEU |o |di|NHEIro [2]

Total cost: 5114.
14 syllables.
2 optimal segmentations.
+400542 possible segmentations (tableau lines).

Frase 4
Marina2-11.wav
#

I: O or · ga · ni · za · dor a · pre · sen · tou a
ca ·
[| ~ ~ ^^^ [~ ^^^ [|
~ ~
ta · lo · ga · do · ra
^^

O: |o |ORganiza|DOR |Apresen|TOU |a |CATaloga|DORa [0]
O: |O organiza|DOR |Apresen|TOU |a |CATaloga|DORa [2]
O: |o |ORganiza|DOR |Apresen|TOU a cataloga|DORa [1]
O: |O organiza|DOR |Apresen|TOU a cataloga|DORa [3]

Total cost: 5015.
17 syllables.
4 optimal segmentations.
+20383750 possible segmentations (tableau lines).

Frase 5
Marina2-11.wav
#

I: A fal · ta de mo · der · ni · za · ção é ca ·
tas
[| ^^^ [| ^^^ [^ | ~ ~
· tró · fi · ca
^^^

O: |a |FALta |DE moderniza|ÇÃO |É catas|TRÓfica [2]
O: |a |FALta de moderniza|ÇÃO |É catas|TRÓfica [1]
O: |a |FALta |de |MODerniza|ÇÃO |É catas|TRÓfica [2]

Total cost: 5014.
15 syllables.
3 optimal segmentations.
+1167180 possible segmentations (tableau lines).

Frase 6
Marina2-11.wav
#

I: O tra . ba . lho da pes . qui . sa . do . ra foi
 [| ^ ^ [| ^ ^ [^ ^ ^
 |
pu . bli . ca . do
 ^ ^

O: |o |tra|BALho |da |PESquisa|DOrA |FOI publi|CAdo [1]
O: |o |tra|BALho |DA pesquisa|DOrA |FOI publi|CAdo [1]

Total cost: 5102.
15 syllables.
2 optimal segmentations.
+2861585 possible segmentations (tableau lines).

Frase 7
Marina2-11.wav
#

I: O pro . fe . ssor tam . bém o des . cre . veu
aos
 [~ | ^ ^ ^ ^ [^ ^ ^ [| ^ ^ ^ [^ ^ ^
 |
a . lu . nos
 ^ ^

O: |o |PROfe|SSOR |tam|BÉM |o |DEScre|VEU |aos |a|LUnos [3]

Total cost: 6203.
14 syllables.
1 optimal segmentation.
+1074186 possible segmentations (tableau lines).

Frase 8

Marina2-11.wav
#

I: O go · ver · na · dor a · cei · tou a mo · der ·
ni ·
[| ^^^ [^^^ [|
za · çã
^^^

O: |O governa|DOR |Acei|TOU a moderniza|ÇÃO [2]
O: |o |GOverna|DOR |Acei|TOU |a |MOderniza|ÇÃO [3]
O: |O governa|DOR |Acei|TOU |a |MOderniza|ÇÃO [3]
O: |o |GOverna|DOR |Acei|TOU a moderniza|ÇÃO [2]

Total cost: 5025.
14 syllables.
4 optimal segmentations.
+765282 possible segmentations (tableau lines).

Frase 9
Marina2-11.wav
#

I: A fal · ta de au · to · ri · da · de foi a ·
lar ·
[| ^^^ [| ~ ~ ^ ^ [^^^ |
man · te
^^^

O: |a |FALta |DE autori|DAde |FOI alar|MANte [2]
O: |a |FALta de autori|DAde |FOI alar|MANte [1]
O: |a |FALta |de |AURI|DAde |FOI alar|MANte [0]

Total cost: 5002.
14 syllables.
3 optimal segmentations.
+898025 possible segmentations (tableau lines).

Frase 10
Marina2-11.wav
#

I: O in · ves · ti · ga · dor já lhe de · vol · veu
o
[| ~ ~ ^^^ [^ ^ | | ^^^
[|
di · nhei · ro

^^^^
 O: |o |INvestiga|DOR |JÁ lhe devol|VEU |o |di|NHEIro [0]
 O: |O investiga|DOR |JÁ lhe devol|VEU |o |di|NHEIro [2]
 O: |o |INvestiga|DOR |JÁ |lhe |DEvol|VEU |o |di|NHEIro [1]
 O: |O investiga|DOR |JÁ |lhe |DEvol|VEU |o |di|NHEIro [3]

Total cost: 5114.
 15 syllables.
 4 optimal segmentations.
 +1385268 possible segmentations (tableau lines).

 # Frase 11
 # Marina2-11.wav
 #

I: A ca · ta · lo · ga · do · ra com · pre · en · deu
 o tra
 [~ | ^ ^ [^ ^ ^ [
 |
 · ba · lho da pes · qui · sa · do · ra
 ^ ^ [~ ~ | ^ ^

O: |a |Cataloga|Dora |COMpreen|DEU |o |tra|BALho |da
 |PESquisa|Dora [5]
 O: |A cataloga|Dora |COMpreen|DEU |o |tra|BALho |da
 |PESquisa|Dora [3]
 O: |a |Cataloga|Dora |COMpreen|DEU |o |tra|BALho |DA
 pesquisa|Dora [3]
 O: |A cataloga|Dora |COMpreen|DEU |o |tra|BALho |DA
 pesquisa|Dora [1]

Total cost: 6104.
 21 syllables.
 4 optimal segmentations.
 +1810779840 possible segmentations (tableau lines).

 # Frase 12
 # Marina2-11.wav
 #

I: A pro · fe · sso · ra dis · cu · tiu a gra · ma
 · ti · ca
 [~ | ^ ^ ^ [^ ^ ^ [|
 · li · da · de
 ^ ^

O: |a |PROfe|SSOra |DIScu|TIU |a |GRAMaticali|DAde [4]

O: |a |PROfe|SSOra |DIScu|TIU a gramaticali|DAde [3]

Total cost: 5002.

16 syllables.

2 optimal segmentations.

+9475728 possible segmentations (tableau lines).

#

Frase 13

Marina2-11.wav

#

I: A in · te · li · gên · cia da ca · ta · lo · ga ·
do · ra

[| ~ ~ ^ ^ ^ [| ~ ~
^ ^ [

foi de · ter · mi · nan · te

^ ^ ^ | ^ ^ ^

O: |a |INteli|GÊNcia |da |CATaloga|DOra |FOI determi|NANte
[0]

O: |A inteli|GÊNcia |da |CATaloga|DOra |FOI determi|NANte
[2]

O: |a |INteli|GÊNcia |DA cataloga|DOra |FOI determi|NANte
[2]

O: |A inteli|GÊNcia |DA cataloga|DOra |FOI determi|NANte [4]

Total cost: 6003.

19 syllables.

4 optimal segmentations.

+147667381 possible segmentations (tableau lines).

#

Frase 14

Marina2-11.wav

#

I: O in · ves · ti · ga · dor já o · fe · re · ceu
di ·

[| ~ ~ ^ ^ ^ [^ ^ | ^ ^ ^ [
nhei · ro
^ ^ ^ ^

O: |o |INvestiga|DOR |JÁ ofere|CEU |di|NHEIro [0]

O: |O investiga|DOR |JÁ ofere|CEU |di|NHEIro [2]

Total cost: 5114.

14 syllables.

2 optimal segmentations.

+302214 possible segmentations (tableau lines).

Frase 15
Marina2-11.wav
#

I: A pro · cu · ra da gra · ma · ti · ca · li · da ·
de é

[~ | ^ [| ^

o no · sso ob · jec · ti · vo
| ^ [^

O: |a |pro|CUra |da |GRAMaticali|DAde |É |o |NOsso
|OBjec|TIvo [3]

O: |a |pro|CUra |DA gramaticali|DAde |É |o |NOsso
|OBjec|TIvo [3]

Total cost: 5102.

20 syllables.

2 optimal segmentations.

+396249028 possible segmentations (tableau lines).

Frase 16
Marina2-11.wav
#

I: A pes · qui · sa · do · ra per · deu au · to · ri
· da · de

[~ | ^ [^ [~

O: |a |PESquisa|DOra |per|DEU |AUTori|DAde [2]

O: |A pesquisa|DOra |per|DEU |AUTori|DAde [0]

Total cost: 4113.

13 syllables.

2 optimal segmentations.

+391005 possible segmentations (tableau lines).

Frase 17
Marina2-11.wav
#

I: O pro · fe · ssor tam · bém des · cre · veu os
a ·

[| ^^^^ [^^^ [^^^ [|
lu · nos
^^

O: |o |PROfe|SSOR |tam|BÉM |DEScRe|VEU |os |a|LUnos [2]

Total cost: 5213.

13 syllables.

1 optimal segmentation.

+312469 possible segmentations (tableau lines).

#

Frase 18

Marina2-11.wav

#

I: A au · to · ri · da · de ca · be ao go · ver ·
na · dor

[~ | ^ ^ [^ ^ [|
^^^

O: |a |AUtori|DAde |CABe ao governa|DOR [2]

O: |a |AUtori|DAde |CABe |ao |GOverna|DOR [3]

O: |a |AUtori|DAde |CABe |AO governa|DOR [3]

O: |A autori|DAde |CABe |AO governa|DOR [1]

O: |A autori|DAde |CABe ao governa|DOR [0]

O: |A autori|DAde |CABe |ao |GOverna|DOR [1]

Total cost: 4013.

13 syllables.

6 optimal segmentations.

+299349 possible segmentations (tableau lines).

#

Frase 19

Marina2-11.wav

#

I: O in · ves · ti · ga · dor já me o · fe · re ·
ceu di

[| ~ ~ ^^^ [^ ^ | |
^^^ [

· nhei · ro

^^^^

O: |o |INvestiga|DOR |JÁ me ofere|CEU |di|NHEIro [0]

O: |O investiga|DOR |JÁ me ofere|CEU |di|NHEIro [2]

Total cost: 5114.

15 syllables.

2 optimal segmentations.
+991233 possible segmentations (tableau lines).

Frase 20
Marina2-11.wav
#

I: A gra · ma · ti · ca · li · da · de das fra · ses
foi

[| ~~~ ^^ [| ^^^
[^^^ |
con · se · gui · da
^^^

O: |a |GRAMaticali|DAde |das |FRAses |FOI conse|GUIDa [0]
O: |A gramaticali|DAde |das |FRAses |FOI conse|GUIDa [2]

Total cost: 5002.
16 syllables.
2 optimal segmentations.
+5559320 possible segmentations (tableau lines).

Test 2

[frases] > ../sotaq --count --rank
**depst%rightmost%clashint%intlex%bingrad%parse:nolapse%clashex
t%trochee%align longo-pb**

Active constraints, absolute weights, relative weights:

| | |
|--------------|-----------|
| align: | 1 |
| bin grad: | 10000 |
| dep st: | 100000000 |
| lex integr: | 100000 |
| nolapse: | 1000 |
| parse: | 1000 |
| rightmost: | 10000000 |
| trochee: | 10 |
| ext noclash: | 100 |
| int noclash: | 1000000 |

#

I: De al · gum mo · do o in · ves · ti · ga · dor
(ou

[| ^^^ [^^ [| + ^^^ [
|

de · ci · sor) pro · cu · ra i · den · ti · fi · car
quais os
^^^^ [^^ | ^^^ [

[
e · ven · tos que são mais pro · vá · veis do
que ou
| ^^^ [| ^^^ | ^^^^ | ^^ [|
| ^^
· tros e às ve · zes vai a · té ao pon · to
de quan
[| | ^^ + [^^^ | ^^ [| ^^^ [

|
· ti · fi · car a pro · ba · bi · li · da · de de
cer · tos
^^^^ [| ^^ [|

^^^^ |
e · ven · tos.
^^^^

O: |de |alGUM |MOdo |o |INvesti|gaDOR |(ou |de|ciSOR)
|pro|CUra
|Identi|fiCAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |aTÉ |ao |PONto |de
|QUANti|fiCAR |a
|PROba|Bili|DAdE |de |CERTos |e|VENTos. [4]

Total cost: 35282.
66 syllables.
1 optimal segmentation.
+19493099733448665395948468592 possible segmentations
(tableau lines).

#

I: Em qual · quer ca · so há sem · pre a pre ·
o · cu ·
[| ^^^ | [^^ [^^^ [| ~~~
pa · ção de pro · cu · rar u · ma "ba · se de a ·
va · li
^^^^ [| ~~~ ^^^ [^ [^^^ + | |
~~
· a · ção" para a pro · ba · bi · li · da · de, quer
se ·
^^^^ [| | ~~~ ~ ^ + [^^^
| ^^

ja com . pa . ra . ti . va, quer se . ja nu . mé . ri
· ca.

| ~ ~ ^ ^ [^ ^ ^ ^ | ^ ^ | ^ ^

O: |em |qualQUER |ca|so HÁ |SEMpre |a |pre|Ocu|paÇÃO |de
|pro|cuRAR
|Uma |"BAsE |de |a|VALi|aÇÃO" |para |a |PROba|Bili|DAde,
|QUER |SEja
|com|PARa|TIva, |QUER |SEja |nu|MÉRica. [3]

Total cost: 136465.

51 syllables.

1 optimal segmentation.

+551060308851130368000 possible segmentations (tableau
lines).

#

I: Os as . pec . tos se . mân . ti . cos são . a . de
· qua

[| ^ ^ ^ | ^ ^ ^ + [^ ^ ^ ~
~ ~ ~

· da . men . te dis . cu . ti . dos em ter . mos da
ba . se

^ ^ ^ [~ ~ ~ ^ ^ [| ^ ^ ^ [|
^ ^ +

de a . va . li . a . ção por ser es . ta que
de . ter

| + | ~ ~ ~ ^ ^ ^ [| ^ ^ ^ | ^ ^ [|
~ ~

· mi . na, pe . lo me . nos em gran . de par . te,
a in

^ ^ [| ^ ^ [| ^ ^ ^ ^ + | ^ ^ ^ +
[|

· ter . pre . ta . ção do con . cei . to de pro . ba
· bi .

~ ~ ~ ^ ^ ^ [| ^ ^ ^ [| ~ ~ ~
~ ~

li . da . de.

^ ^ +

O: |os |as|PECTos |se|MÂNTicos |SÃOade|QUAda|MENTe
|DIScu|TIDOS |em

|TERmos |da |BAsE |de |a|VALi|aÇÃO |por |SER |ESTa |que
|DEter|MINa,

|pe|lo MENos |em |GRANde |PARte, |a |in|TERpre|taÇÃO |do
|con|CEito

|de |PROba|Bili|DAde. [4]

Total cost: 136152.
 66 syllables.
 1 optimal segmentation.
 +66736757338891581982459622616 possible segmentations
 (tableau lines).

 #

I: Re · co · nhe · cem- · se, co · rre · ta · men · te,
 as se
 [~ ~ ^ ^ [~ ~ ^ ^ + [|
 |
 · guin · tes ba · ses de a · va · li · a · çãõ:
 ^ ^ ^ ^ + | ^ ^ | + | ~ ~ ^ ^ ^ ^
 O: |REco|NHEcem-se, |CORre|taMENTe, |as |seGUINTes |BASes
 |de |a|VALi|aÇÃO: [2]

Total cost: 11136.
 22 syllables.
 1 optimal segmentation.
 +3201674599 possible segmentations (tableau lines).

#

[frases] > ../sotaq --count --rank
depst:rightmost:clashint%trochee%align:intlex%parse%clashext%
footbin:nolapse longo-pe

Active constraints, absolute weights, relative weights:

| | |
|--------------|--------|
| align: | 1000 |
| dep st: | 100000 |
| binary foot: | 1 |
| lex integr: | 1000 |
| nolapse: | 1 |
| parse: | 100 |
| rightmost: | 100000 |
| trochee: | 10000 |
| ext noclash: | 10 |
| int noclash: | 100000 |

#

I: De al · gum mo · do o in · ves · ti · ga · dor
 (ou
 [| ^ ^ [^ ^ [| + ^ ^ [|
 |
 de · ci · sor) pro · cu · ra i · den · ti · fi · car
 quais os

e · ven · tos que são mais pro · vá · veis do
 que ou
 · tros e às ve · zes vai a · té ao pon · to
 de quan
 · ti · fi · car a pro · ba · bi · li · da · de de
 cer · tos
 e · ven · tos.

O: |de |al|GUM |MOdo o investiga|DOR |(ou |DEci|SOR)
 |pro|CUra
 |Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
 |pro|VÁveis |do
 |que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto de
 quantifi|CAR a
 probabili|DAde |de |CERTos |e|VENTos. [2]

O: |de |al|GUM |MOdo o investiga|DOR |(ou |DEci|SOR)
 |pro|CUra
 |Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
 |pro|VÁveis |do
 |que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |DE
 quantifi|CAR |a
 |PRObabili|DAde |de |CERTos |e|VENTos. [4]

O: |de |al|GUM |MOdo o investiga|DOR |(ou |DEci|SOR)
 |pro|CUra
 |Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
 |pro|VÁveis |do
 |que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |de
 |QUANTifi|CAR |a
 |PRObabili|DAde |de |CERTos |e|VENTos. [4]

O: |de |al|GUM |MOdo o investiga|DOR |(ou |DEci|SOR)
 |pro|CUra
 |Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
 |pro|VÁveis |do
 |que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |de
 |QUANTifi|CAR a
 probabili|DAde |de |CERTos |e|VENTos. [3]

O: |de |al|GUM |MOdo |O investiga|DOR |(ou |DEci|SOR)
 |pro|CUra

|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |de
|QUANTifi|CAR a
probabili|DAde |de |CERTos |e|VENTos. [4]
O: |de |al|GUM |MODO |o |INvestiga|DOR |(ou |DEci|SOR)
|pro|CUra
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto de
quantifi|CAR |a
|PRObabili|DAde |de |CERTos |e|VENTos. [4]
O: |de |al|GUM |MODO |o |INvestiga|DOR |(ou |DEci|SOR)
|pro|CUra
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto de
quantifi|CAR a
probabili|DAde |de |CERTos |e|VENTos. [3]
O: |de |al|GUM |MODO |o |INvestiga|DOR |(ou |DEci|SOR)
|pro|CUra
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |de
|QUANTifi|CAR |a
|PRObabili|DAde |de |CERTos |e|VENTos. [5]
O: |de |al|GUM |MODO |O investiga|DOR |(ou |DEci|SOR)
|pro|CUra
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto de
quantifi|CAR a
probabili|DAde |de |CERTos |e|VENTos. [3]
O: |de |al|GUM |MODO |o |INvestiga|DOR |(ou |DEci|SOR)
|pro|CUra
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |DE
quantifi|CAR |a
|PRObabili|DAde |de |CERTos |e|VENTos. [5]
O: |de |al|GUM |MODO |o investiga|DOR |(ou |DEci|SOR)
|pro|CUra
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto de
quantifi|CAR |a
|PRObabili|DAde |de |CERTos |e|VENTos. [3]

O: |de |al|GUM |MODO |O investiga|DOR |(ou |DEci|SOR)
|pro|Cura
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto de
quantifi|CAR |a
|PRObabili|DAde |de |CERTos |e|VENTos. [4]
O: |de |al|GUM |MODO |O |INvestiga|DOR |(ou |DEci|SOR)
|pro|Cura
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |DE
quantifi|CAR |a
probabili|DAde |de |CERTos |e|VENTos. [4]
O: |de |al|GUM |MODO |O |INvestiga|DOR |(ou |DEci|SOR)
|pro|Cura
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |de
|QUANTifi|CAR |a
probabili|DAde |de |CERTos |e|VENTos. [4]
O: |de |al|GUM |MODO |O investiga|DOR |(ou |DEci|SOR)
|pro|Cura
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |DE
quantifi|CAR |a
probabili|DAde |de |CERTos |e|VENTos. [3]
O: |de |al|GUM |MODO |O investiga|DOR |(ou |DEci|SOR)
|pro|Cura
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |de
|QUANTifi|CAR |a
|PRObabili|DAde |de |CERTos |e|VENTos. [5]
O: |de |al|GUM |MODO |O investiga|DOR |(ou |DEci|SOR)
|pro|Cura
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do
|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |DE
quantifi|CAR |a
|PRObabili|DAde |de |CERTos |e|VENTos. [5]
O: |de |al|GUM |MODO |O investiga|DOR |(ou |DEci|SOR)
|pro|Cura
|Identifi|CAR |quais |os |e|VENTos |que |SÃO |MAIS
|pro|VÁveis |do

|que |OUTros |e |às |VEzes |VAI |a|TÉ |ao |PONto |DE
quantifi|CAR a
probabili|DAde |de |CERTos |e|VENTos. [4]

Total cost: 22633.

66 syllables.

18 optimal segmentations.

+19493099733448665395948468592 possible segmentations
(tableau lines).

#

I: Em qual · quer ca · so há sem · pre a pre ·
o · cu ·
[| ^^^^ | [^^ [^^^ | |
pa · ção de pro · cu · rar u · ma "ba · se de a ·
va · li
^^^ [| ^^^ [^ [^^^ + | |
· a · ção" para a pro · ba · bi · li · da · de, quer
se ·
| ^^^^ [| | ^ + [^^^^
| ^^
ja com · pa · ra · ti · va, quer se · ja nu · mé · ri
· ca.

O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
|PROcu|RAR |Uma
|"BASE de avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER |SEja
compara|TIva, |QUER |SEja |nu|MÉRica. [2]

O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
|PROcu|RAR |Uma
|"BASE |de |Avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER |SEja
compara|TIva, |QUER |SEja |nu|MÉRica. [3]

O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
|PROcu|RAR |Uma
|"BASE |de |Avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
compara|TIva, |QUER |SEja |nu|MÉRica. [2]

O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BASE |de |Avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
|SEja
compara|TIva, |QUER |SEja |nu|MÉRica. [4]

O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BASE de avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
|SEja
|COMpara|TIva, |QUER |SEja |nu|MÉRica. [4]

O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BAsE de avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
|COMpara|TIva, |QUER |SEja |nu|MÉrica. [3]

O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
|PROcu|RAR |Uma
|"BAsE de avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
|COMpara|TIva, |QUER |SEja |nu|MÉrica. [2]

O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BAsE |de |Avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
|SEja
|COMpara|TIva, |QUER |SEja |nu|MÉrica. [5]

O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
|PROcu|RAR |Uma
|"BAsE de avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
compara|TIva,
|QUER |SEja |nu|MÉrica. [1]

O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BAsE de avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
|COMpara|TIva, |QUER |SEja |nu|MÉrica. [3]

O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
|PROcu|RAR |Uma
|"BAsE |de |Avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER |SEja
|COMpara|TIva, |QUER |SEja |nu|MÉrica. [4]

O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BAsE |de |Avalia|ÇÃO" para a probabili|DAde, |QUER
|SEja
|COMpara|TIva, |QUER |SEja |nu|MÉrica. [4]

O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BAsE de avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
compara|TIva, |QUER |SEja |nu|MÉrica. [2]

O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BAsE |de |Avalia|ÇÃO" para a probabili|DAde, |QUER
|SEja
|COMpara|TIva, |QUER |SEja |nu|MÉrica. [4]

O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
|PROcu|RAR
|Uma |"BAsE |de |Avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
|SEja
compara|TIva, |QUER |SEja |nu|MÉrica. [4]

O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
|PROcu|RAR |Uma

|"BAsE de avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER |SEja
 |COMpara|TIva, |QUER |SEja |nu|MÉrica. [3]
 O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
 |PROcu|RAR
 |Uma |"BAsE de avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
 |SEja
 compara|TIva, |QUER |SEja |nu|MÉrica. [3]
 O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
 |PROcu|RAR
 |Uma |"BAsE de avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
 compara|TIva, |QUER |SEja |nu|MÉrica. [2]
 O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
 |PROcu|RAR
 |Uma |"BAsE de avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
 |SEja
 compara|TIva, |QUER |SEja |nu|MÉrica. [3]
 O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
 |PROcu|RAR
 |Uma |"BAsE |de |Avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
 |SEja
 |COMpara|TIva, |QUER |SEja |nu|MÉrica. [5]
 O: |em |qual|QUER caso |HÁ |SEMpre a preocupa|ÇÃO |de
 |PROcu|RAR |Uma
 |"BAsE |de |Avalia|ÇÃO" para a probabili|DAde, |QUER |SEja
 |COMpara|TIva, |QUER |SEja |nu|MÉrica. [3]
 O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
 |PROcu|RAR
 |Uma |"BAsE de avalia|ÇÃO" |para |a |PRObabili|DAde, |QUER
 |SEja
 |COMpara|TIva, |QUER |SEja |nu|MÉrica. [4]
 O: |em |qual|QUER caso |HÁ |SEMpre |A preocupa|ÇÃO |de
 |PROcu|RAR
 |Uma |"BAsE |de |Avalia|ÇÃO" para a probabili|DAde, |QUER
 |SEja
 compara|TIva, |QUER |SEja |nu|MÉrica. [3]
 O: |em |qual|QUER caso |HÁ |SEMpre |a |PREocupa|ÇÃO |de
 |PROcu|RAR
 |Uma |"BAsE |de |Avalia|ÇÃO" para a probabili|DAde, |QUER
 |SEja
 compara|TIva, |QUER |SEja |nu|MÉrica. [3]

Total cost: 15252.

51 syllables.

24 optimal segmentations.

+551060308851130368000 possible segmentations (tableau lines).

I: Os as · pec · tos se · mân · ti · cos são · a · de
· qua
[| ^^^ | ^^^ + [^^^
· da · men · te dis · cu · ti · dos em ter · mos da
ba · se
^^^ [^^ [| ^^^ [|
^^ +
de a · va · li · a · ção por ser es · ta que
de · ter
| + | ^^^ [| ^^^ | ^^ [|
· mi · na, pe · lo me · nos em gran · de par · te,
a in
^^ [| ^^ [| ^^^^ + | ^^^ +
[|
· ter · pre · ta · ção do con · cei · to de pro · ba
· bi ·
^^^ [| ^^^ [|
li · da · de.
^^ +

O: |os |as|PECTos |se|MÂNTicos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BAsE de avalia|ÇÃO |por |SER |ESta |que
|DEter|MIIna, pelo
|MENos |em |GRANde |PARte, |a |INterpreta|ÇÃO |do |con|CEIto
|de
|PRObabili|DAdE. [4]

O: |os |as|PECTos |se|MÂNTicos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BAsE |de |Avalia|ÇÃO |por |SER |ESta |que
|DEter|MIIna,
pelo |MENos |em |GRANde |PARte, |a |INterpreta|ÇÃO |do
|con|CEIto |DE
probabili|DAdE. [5]

O: |os |as|PECTos |se|MÂNTicos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BAsE de avalia|ÇÃO |por |SER |ESta |que
|DEter|MIIna, pelo
|MENos |em |GRANde |PARte, |A interpreta|ÇÃO |do |con|CEIto
|DE
probabili|DAdE. [4]

O: |os |as|PECTos |se|MÂNTicos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BAsE |de |Avalia|ÇÃO |por |SER |ESta |que
|DEter|MIIna,

pelo |MENoS |em |GRANde |PARte, |a |INterpreta|ÇÃO |do
|con|CEito |de
|PRObabili|DAde. [5]

O: |os |as|PECTos |se|MÂnticos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BASE de avalia|ÇÃO |por |SER |ESTa |que
|DEter|MIIna, pelo
|MENoS |em |GRANde |PARte, a interpreta|ÇÃO |do |con|CEito
|de
|PRObabili|DAde. [3]

O: |os |as|PECTos |se|MÂnticos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BASE de avalia|ÇÃO |por |SER |ESTa |que
|DEter|MIIna, pelo
|MENoS |em |GRANde |PARte, |A interpreta|ÇÃO |do |con|CEito
|de
|PRObabili|DAde. [4]

O: |os |as|PECTos |se|MÂnticos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BASE |de |Avalia|ÇÃO |por |SER |ESTa |que
|DEter|MIIna,
pelo |MENoS |em |GRANde |PARte, a interpreta|ÇÃO |do
|con|CEito |DE
probabili|DAde. [4]

O: |os |as|PECTos |se|MÂnticos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BASE de avalia|ÇÃO |por |SER |ESTa |que
|DEter|MIIna, pelo
|MENoS |em |GRANde |PARte, a interpreta|ÇÃO |do |con|CEito
|DE
probabili|DAde. [3]

O: |os |as|PECTos |se|MÂnticos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BASE de avalia|ÇÃO |por |SER |ESTa |que
|DEter|MIIna, pelo
|MENoS |em |GRANde |PARte, |a |INterpreta|ÇÃO |do |con|CEito
|DE
probabili|DAde. [4]

O: |os |as|PECTos |se|MÂnticos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BASE |de |Avalia|ÇÃO |por |SER |ESTa |que
|DEter|MIIna,
pelo |MENoS |em |GRANde |PARte, |A interpreta|ÇÃO |do
|con|CEito |DE
probabili|DAde. [5]

O: |os |as|PECTos |se|MÂnticos |SÃO adequada|MENTe
|DIScu|TIdos |em

|TERmos |da |BAse |de |Avalia|ÇÃO |por |SER |ESta |que
|DEter|MIIna,
pelo |MEnos |em |GRANde |PARte, a interpreta|ÇÃO |do
|con|CEIto |de
|PRObabili|DAde. [4]
O: |os |as|PECTos |se|MÂNTicos |SÃO adequada|MENTe
|DIScu|TIdos |em
|TERmos |da |BAse |de |Avalia|ÇÃO |por |SER |ESta |que
|DEter|MIIna,
pelo |MEnos |em |GRANde |PARte, |A interpreta|ÇÃO |do
|con|CEIto |de
|PRObabili|DAde. [5]

Total cost: 22318.
66 syllables.
12 optimal segmentations.
+77875765223140411886380027320 possible segmentations
(tableau lines).

I: Re · co · nhe · cem- · se, co · rre · ta · men · te,
as se
[^^^ [^^^ +]
|
· guin · tes ba · ses de a · va · li · a · çãõ:
 ^^^^ + ^^ | + | ^^^^^
O: |REco|NHEcem-se, |CORreta|MENTe, |as |se|GUINtes |BASes
|de |Avalia|ÇÃO: [3]
O: |REco|NHEcem-se, |CORreta|MENTe, |as |se|GUINtes |BASes
de avalia|ÇÃO: [2]

Total cost: 6114.
22 syllables.
2 optimal segmentations.
+3201674599 possible segmentations (tableau lines).

Test 3

```
[frases] > ../sotaq --count --rank
depst:rightmost:clashint%intlex%bingrad:parse:nolapse%clashex
t%trochee%align long
Active constraints, absolute weights, relative weights:
  align:          1
  bin grad:      1000
  dep st:        100000
```

lex integr: 10000
 nolapse: 1000
 parse: 1000
 rightmost: 100000
 trochee: 10
 ext noclash: 100
 int noclash: 100000

I: A gra · ma · ti · ca · li · da · de foi com · pu ·
 ta ·
 | | ^ ^ | ^ ^ ^ |
 do · ri · za · da.
 ^ ^

O: |a|gra|MAti|CAli|DAde|FOI|com|PUta|DORi|ZAda. [4]
 O: |a|GRAMati|CALi|DAde|FOI|com|PUta|DORi|ZAda. [4]

Total cost: 3007.
 16 syllables.
 2 optimal segmentations.
 +5146083 possible segmentations (tableau lines).

I: A gra · ma · ti · ca · li · da · de foi com · pu ·
 ta ·
 | | ^ ^ | |
 do · ri · za · da.
 ^ ^

O: |a|GRAMati|CALi|DAde|foi|com|PUta|DORi|ZAda. [4]
 O: |a|gra|MAti|CALi|DAde|foi|COMputa|DORi|ZAda. [4]
 O: |a|GRAMati|CALi|DAde|foi|COMputa|DORi|ZAda. [4]
 O: |a|gra|MAti|CALi|DAde|foi|com|PUta|DORi|ZAda. [4]

Total cost: 2006.
 16 syllables.
 4 optimal segmentations.
 +8517653 possible segmentations (tableau lines).

I: O cons · tan · ti · no · po · lis · mo cons · tan ·
 ti · no
 | | ^ ^ ^ |
 · po · lis · ta cons · tan · ti · no · po · li · zou a
 ca · ta
 ^ ^ ^ | ^ ^ ^ | |
 · lo · ga · do · ra.
 ^ ^

O: |o|CONStanti|NOpo|LISmo|cons|TANti|NOpo|LISta
|CONStan|TIIno|
Poli|ZOU|a|CAta|LOga|DOra.[9]
O: |o|cons|TANti|NOpo|LISmo|CONStanti|NOpo|LISta
|CONStan|TIIno|
Poli|ZOU|a|CAta|LOga|DOra.[9]
O: |o|cons|TANti|NOpo|LISmo|cons|TANti|NOpo|LISta
|CONStan|TIIno|
Poli|ZOU|a|CAta|LOga|DOra.[9]
O: |o|CONStanti|NOpo|LISmo|CONStanti|NOpo|LISta
|CONStan|TIIno|
Poli|ZOU|a|CAta|LOga|DOra.[9]

Total cost: 3013.
29 syllables.
4 optimal segmentations.
+3754477858148 possible segmentations (tableau lines).

I: A Fla · vi · a · ne ca · ta · lo · ga · ri · a I ·
ta · qua
| | ^ | ^^ |
· que · ce · tu · ba.
^^

O: |a|FLAvi|Ane|CAta|LOga|RIa|i|TAqua|QUEce|TUba.[5]
O: |a|FLAvi|Ane|CAta|LOga|RIa|Itaqua|QUEce|TUba.[5]

Total cost: 1008.
18 syllables.
2 optimal segmentations.
+67442853 possible segmentations (tableau lines).

[frases] > ../sotaq --count --rank
depst:rightmost:clashint%trochee%align:intlex%parse%clashext%
footbin:nolapse long

Active constraints, absolute weights, relative weights:

align: 1000
dep st: 100000
binary foot: 1
lex integr: 1000
nolapse: 1
parse: 100
rightmost: 100000
trochee: 10000
ext noclash: 10
int noclash: 100000

I: A gra · ma · ti · ca · li · da · de foi com · pu ·
 ta · do
 | | ^ ^ | ^ ^ ^ |
 · ri · za · da.
 ^ ^

O: |a |GRAmaticali|DAde |FOI computadori|ZAda. [1]

Total cost: 5002.
 16 syllables.
 1 optimal segmentation.
 +5146083 possible segmentations (tableau lines).

I: A gra · ma · ti · ca · li · da · de foi com · pu ·
 ta · do
 | | ^ ^ | | |
 · ri · za · da.
 ^ ^

O: |a |GRAmaticali|DAde |foi |COMputadori|ZAda. [2]

Total cost: 4002.
 16 syllables.
 1 optimal segmentation.
 +8517653 possible segmentations (tableau lines).

I: O cons · tan · ti · no · po · lis · mo cons · tan ·
 ti · no
 | | ^ ^ ^ | |
 · po · lis · ta cons · tan · ti · no · po · li · zou a
 ca · ta

· lo · ga · do · ra.
 ^ ^ ^ | ^ ^ ^ | |

O: |o |CONStantinopo|LISmo |CONStantinopo|LISTa
 |CONStantinopoli|ZOU
 a cataloga|DOra. [3]

O: |o |CONStantinopo|LISmo |CONStantinopo|LISTa
 |CONStantinopoli|ZOU
 |a |CATaloga|DOra. [4]

Total cost: 8005.
 29 syllables.
 2 optimal segmentations.
 +3754477858148 possible segmentations (tableau lines).

I: A Fla · vi · a · ne ca · ta · lo · ga · ri · a I ·
 ta · qua
 | | ^ | ^^ |
 · que · ce · tu · ba.
 ^^

O: |a |FLAvi|Ane |CAataloga|RIa |Itaquaquece|TUba. [3]

Total cost: 6002.

18 syllables.

1 optimal segmentation.

+67442853 possible segmentations (tableau lines).

References:

- Abaurre, M.B. 1979. Phonostylistic aspects of a Brazilian Portuguese dialect: implications for syllable-structure constraints. SUNY/Buffalo Ph.D. dissertation.
- Abaurre, M.B. & C. Galves. 1998. Rhythmic differences between European and Brazilian Portuguese: an optimalist and minimalist approach. *DELTA* 14: 377-403.
- Bisol, L. 1992. O acento: Duas alternativas de análise. Ms., Pontifícia Universidade Católica do Rio Grande do Sul, Brazil.
- _____. 2000. O troqueu silábico no sistema fonológico: um adendo ao artigo de Plínio Barbosa. *DELTA* 16: 403-413.
- Camara Jr., Mattoso. 1953. *Para o estudo da fonemica portuguesa*. Rio de Janeiro: Organização Simões.
- Carvalho, J. B. 1988. Réduction vocalique, quantité et accentuation: pour une explication structurale de la divergence entre portugais lusitanien et portugais brésilien. *Boletim de Filologia* 32:5-26.
- _____. 1989. Phonological conditions on portuguese clitic placement: on syntactic evidence for stress and rhythmical patterns. *Linguistics* 27: 405-436.
- Collischonn, G. 1993. *Um estudo do acento secundário em português*. Universidade Federal do Rio Grande do Sul master thesis.
- D'Andrade, E. & S.B. Laks. 1991. Na crista da onda: o acento de palavra em português. *Actas do 7º Encontro da Associação Portuguesa de Linguística*: 15-26.
- Frota, S. 1998. *Prosody and Focus in European Portuguese*. Universidade de Lisboa Ph.D. Dissertation.
- Harris, J. W. 1989. The Stress Erasure Convention and Cliticization in Spanish. *Linguistic Inquiry* 20: 339-363.
- Hayes, B. 1995. *Metrical stress theory: Principles and Case Studies*. University of Chicago Press.
- Hulst, H. V. D. 1997. Primary accent is non-metrical. *Rivista di Linguistica* 8.1 [guest editor: Irene Vogel].
- Lee, S. H. 1984. A regra de acento do português: outra alternativa. *Letras de Hoje*, 37-42.
- Kager, R. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

- Massini-Cagliari, G. 1995. *Cantigas de amigo; do ritmo poético ao lingüístico – um percurso histórico da acentuação em português*. Universidade Estadual de Campinas Ph.D. dissertation.
- Mateus, M. H. M. 1975. *Aspectos da Fonologia do Português*. Lisboa: Instituto Nacional de Investigação Científica.
- Mateus, M. H. M. et alii. 1983. *Gramática da Língua Portuguesa*. Coimbra: Almedina.
- Prince, A. & P. Smolensky. 1993. Optimality Theory. Rutgers Optimality Archive.
- Prince, A. 1983. Relating to the Grid. *Linguistic Inquiry* 14: 19-100.
- Reynolds, W. T. 1994. Variation and phonological theory. University of Pennsylvania Ph.D. dissertation.
- Roca, I. M. 1986. Secondary Stress and Metrical Rhythm. *Phonology Yearbook* 3: 341-370.
- Sandalo, F. & M. B. Abaurre & C. Galves 1999. Otimizando o ritmo do português. Paper presented at III CELSUL, Pontifícia Universidade Católica do Rio Grande do Sul, Brazil.
- Sandalo, F. & H. Truckenbrodt. 2001. Some Notes on Phonological Phrasing in Brazilian Portuguese. To appear in MIT Working Papers in Linguistics.
- Vigário, M. 1998. Cliticização no português europeu: uma operação pós-lexical. XIV Encontro Nacional da Associação Portuguesa de Lingüística, Aveiro, Portugal.