A Grammar to Voice Output Applications in Portuguese Language *

Alex Maia Soares¹, Mirocem Fernandes de Oliveira², Carlos Michel Betemps³

¹Tumelero S/A
Porto Alegre – RS – Brazil

² Adall Industrial Eletrônica Ltda
Porto Alegre – RS – Brazil

³Universidade Federal do Pampa – UNIPAMPA
Campus Bagé (UFPe) – Bagé, RS – Brazil

maiares@maiares.com.br, mirocem2@yahoo.com.br, carlos.betemps@gmail.com

* This work was realized in the Centro Universitário La Salle - UNILASALLE
Canoas, RS – Brazil

Abstract. The communication and the access to information has been a constant concern of the human being since of civilization early days. In this area, the people with visual deficiency have his/her part of interest. The Braille method was one of the first attempts to make better the "relationship" of the visual deficient people with the information. In this work is presented a grammar to processing portuguese texts as input and make an spoken output to these texts. This grammar provides a base to build voice output applications. This type of application is useful mainly to people which have some visual deficiency. Some other aspects like treatment of diphthongs ("ditongos" in portuguese), hiatus ("hiatos" in portuguese), prefixes (DES, DIS, TRANS, CIS, SUB, SUPER, HIPER, INTER, AB, BIS), consonant 's' between vowels, pseudo hiatus, and distinction between diphthongs and hiatus by context (near letters) are too presented.

Resumo. A comunicação e o acesso à informação têm sido uma constante preocupação dos ser humano desde o início da civilização. Nesta área, as pessoas com deficiência visual têm grande interesse. O método Braille foi uma das primeiras tentativas de tornar melhor a "relação" das pessoas com deficiência visual com a informação. Neste trabalho é apresentada uma gramática para o processamento de textos na língua portuguesa e produção de uma saída "falada" destes textos. Esta gramática fornece uma base para a construção de aplicativos de saída de voz. Este tipo de aplicativo é útil, principalmente, para pessoas que possuam alguma deficiência visual. Outros aspectos como o tratamento de ditongos, hiatus, prefixos (DES, DIS, TRANS, CIS, SUB, SUPER, HIPER, INTER, AB, BIS), consoante 's' entre vogais, pseudo-hiatos e distinção entre ditongos e hiatus através do contexto (letras próximas) também são apresentados.
1. Introduction

The communication and the access to information has been a constant concern of the human being since of civilization early days. In this area, the people with visual deficiency have his/her part of interest. The Braille method was one of the first attempts to make better the "relationship" of the visual deficient people with the information.

In this work is presented a grammar to processing portuguese texts as input and make an spoken output to these texts, providing a base to build voice output applications. This type of programs is very useful to visual deficient people interact with computers. Some other aspects like treatment of diphthongs ("ditongos" in portuguese), hiatus ("hiatos" in portuguese), special prefixes (DES, DIS, TRANS, CIS, SUB, SUPER, HIPER, INTER, AB, BIS), consonant 's' between vowels, pseudo hiatus, and distinction between diphthongs and hiatus by context (near letters) are too presented.

Voice output applications use an input text (or word) and make the output related to this input hear capable, i.e., produce a voice "speaking" the input text. This voice can be synthesized or previously recorded by a human (according the language phonemes). Some programs in this category, and indicated by a Visual Deficient Association [Adeva, 2005], are: (a) JUNO - provided by Dolphin [Dolphin, 2005]; (b) DOSVOX - is a academic project by the Federal University of Rio de Janeiro [Dosvox, 2005]; (c) Virtual Vision - developed by the MicroPower [MicroPower, 2005].

All this programs need a device to processing the given text as input (obtained in some way). In this point arise the grammars study [Menezes, 2000]. One manner to processing texts (in a given language) is build a grammar and implements this grammar in some programming language. This was the approach used in this paper.

This paper is structured as follow: Section 2 presents the structures of the spoken portuguese language. Next, the grammar to process texts in portuguese language, including accentuation, is presented in Section 3. In Section 4 some aspects of the portuguese language that need of special treatment, after the grammar processing, are described. Section 5 describes the programs developed in this work and presents an experiment with these programs, which were submitted to the processing of 35885 portuguese words. Section 6 summarizes the contribution of the paper and indicates future works.

2. Spoken Structures of Portuguese

To define a grammar to the portuguese idiom, is necessary understand its phonemes. In [Soares, 2004], ten structures of portuguese language were identified. Later [Oliveira, 2005], these structures were improved to handle with some phoneme aspects (mainly accentuation). These structures are described below [Soares, 2004] [Oliveira, 2005] and the complete list of phonemes is presented in [Oliveira, 2005]:

- Structure V - this structure maintains the vowels ('a', 'e', 'i', 'o' and 'u'). Besides, the accentuated vowels are in this structure too.
- Structure VL - combining the vowels with the consonants 'l', 'm', 'n', 'r' and 's' (with the accentuated vowels) the VL set is defined: An example of a phoneme of this structure is in the portuguese word "a.flu.ên.cia" (phoneme in italic and using a dot '.' to separate the syllables).
- Structure CV - the structure CV comes out of a mixing of the consonants 'b', 'c', 'd', 'f', 'g', 'h', 'j', 'l', 'm', 'n', 'p', 'q', 'r', 's', 't', 'v', 'x', 'z' with the vowels, adding the 'ç' and the accentuated vowels, like in the word "am.bi.guo" (phoneme in italic).
- Structure CVL - this is a mix of the structures CV and VL, like as "bál.sa.mo" (phoneme in italic). The phonemes in that structure are demonstrated in [Soares, 2004].

- Structure CVLS - this structure is derived from CVL, but includes the letter 's' in the phoneme's end. The only phoneme in that structure is gens. An example word is "fer.ra.gens" (phoneme in italic).

- Structure CVU - combining the CV and V structures and using the vowel 'u' in the middle of the syllable, the CVU structure is defined. The vowel 'u' is not followed by other 'u'. Example: "ar.qué.ti.po" (phoneme in italic).

- Structure CVUL - the CVU structure mixed with the consonants 'l', 'm', 'n', 'r' and 's' produce the structure CVUL. An example of this structure is the portuguese word "qüin.qua.ge.si.mo" (phoneme in italic).

- Structure CHV - this structure is the union of two consonants with a vowel. The second consonant may be one of the letters 'l', 'm', 'r' or 's', like the word "o.blí.quo" (phoneme in italic).

- Structure CHVL - this structure is obtained from the mix of CHV and VL structures. The word "a.brân.quio" (phoneme in italic) is an example.

- Structure CHVLS - the structure CHVLS is the structure CHVL with the adding of the consonant 's' in the end of the phoneme. However, only the phonemes "trän" and "tren", from the structure CHVL, are used. An example is the word "trâns.fu.ga" (phoneme in italic)

3. A Grammar to the Portuguese Language Structures

After the portuguese language study, a grammar to recognize this phonemes was defined and is presented below. The grammar provides a base to building applications with voice output to the portuguese language. The grammar is presented using the notation in conformity with [Menezes, 2000].

\[
G = (V, T, P, S)
\]

\[
V = \{S, <CHVLS>, <CHVL>, <CVLS>, <CVUL>, <CVU>, <CHV>, <CV>, <VL>, <V>, <CONS>, <VOGAL>, <HRLS>, <MNRSL>, <Ss>, <HRLSN>, <QG>, <Nu>, <AEIO}\}
\]

\[
T = \{A, Á, Æ, E, É, Ê, I, Í, Ò, Ó, Ô, Ù, Ú, a, á, æ, e, é, ê, i, í, o, ó, ô, ù, ú, b, c, ç, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, x, y, z\}
\]

\[
P = \{
S \rightarrow <CVLS> | <CHVL> | <CVLS> | <CVUL> | <CVU> | <CHV> | <CV> | <VL> | <V> | ε

<CHVLS> \rightarrow <CHS> <HRLS> <MNRSL> <Ss> <CONS> S,
<CHVL> \rightarrow <CONS> <HRLS> <VOGAL> <MNRSL> <Ss> S,
<CVLS> \rightarrow <CONS> <VOGAL> <CONS> S,
<CVUL> \rightarrow <QG> <CONS> <CONS> S,
<CVU> \rightarrow <QG> <CONS> <CONS> S,
<CHV> \rightarrow <CONS> <HRLSN> <VOGAL> S,
<CVL> \rightarrow <CONS> <VOGAL> <MNRSL> <CONS> S,
<CV> \rightarrow <CONS> <VOGAL> S,
<VL> \rightarrow <VOGAL> <MNRSL> <CONS> S,
<V> \rightarrow <V> S,
<CONS> \rightarrow [c][d][f][g][h][j][k][l][m][n][p][q][r][s][t][v][x][y][z],
<VOGAL> \rightarrow [a][á][à][e][ê][é][í][í][ó][ö][õ][ô][ú][ü],
<HRLS> \rightarrow [h][l][s][h][r][l][s],
<MNRSL> \rightarrow [m][n][r][s][l],
<CHRLSN> \rightarrow [c][c][c][c][m][n][r][s][l],
<CONS> \rightarrow S,
<QG> \rightarrow Q,q,G,g,
\}

Hifen, Uruguaiana, V. 30, n° 58, II semestre 2006
Where: G: is the grammar; V: set of variable symbols (not terminals); T: set of terminals symbols; P: set of production rules; S: start symbol; \( \varepsilon \): blank sentence

In the implementation of the grammar, the analysis of the phonemes in any text must follow the next approach. The words are picked one by one (the words are separated by a blank space in a text). Next, an attempting to recognize a phoneme (the structures described in Sect. 2) in that word is made with five letters (the biggest phoneme identified have 5 letters) of the word. If the grammar recognizes that letters set as one of the defined structures, then the text is spoke (using a synthesized or a recorded voice) and the next phoneme will be analyzed using the same process. On contrario, the five letters is not a phoneme, and then four letters are picked and analyzed using the same process. The analysis goes on with three letters, two, and finally with one letter. The process stops when a phoneme is found. Resuming, the analysis must occur from the biggest phoneme to the smallest one for guarantee the correct phoneme recognition.

4. Enhancing the Grammar

There is the necessity to enhance the grammar (indeed the implementation of the grammar) to handle with some aspects of the portuguese language, mainly in analysis of:

- Diphthongs ("ditongos" in portuguese). A diphthong is the junction of a vowel and a semivowels, or vice-versa, in the same syllable. A diphthong must be not separated in two syllables [Vasselai, 2003];
- Hiatus ("hiatos" in portuguese). A hiatus is the sequence of two vowels in separated syllables. A hiatus must be separated in two syllables [Vasselai, 2003];
- Special prefixes: DES, DIS, TRANS, CIS, SUB, SUPER, HIPER, INTER, AB, BIS;
- Consonant 's' between vowels;
- Pseudo Hiatus;
- Distinction between diphthongs and hiatus by context;

These aspects are described below [Oliveira, 2005]:

- Open diphthongs - The open diphthongs ("ditongos abertos" in portuguese) "ÉU", "ÉI", "ÓI" and more the related consonants form a syllable. This must not be separated. Examples (using a dot to separate the syllables): "ES.CAR.C É U", "TRO.F É U", "AN.D R Ó I.DE" (diphthongs in italic).
- Special prefixes - In words with the prefixes DES, DIS, TRANS, CIS, SUB, SUPER, HIPER, INTER, AB, and BIS, and if these are followed by a vowel, then the last consonant of the prefix must be in the next syllable. This prefix processing must be done to enhance the phonetic pronounce. Example:
  - "DES.EN.TEN.DER" must be "DE.SEN.TEN.DER";
  - "IN.TER.ES.TA.DU.AL" must be "IN.TE.RES.TA.DU.AL".
- Consonant 's' - The consonant 's' between two vowels have the same "sound" of the consonant 'z'. Thus, always that the letter 's' is between two vowels that letter will be replaced by the letter 'z' (only for a phonetic reproduction).
- Pseudo Hiatus - Pseudo hiatus occur when the open diphthongs ("ditongos abertos" in portuguese) "ÉI" and "ÓI" are followed by the vowel 'a'. This combination must not be

- Distinction between diphthongs and hiatus by the context - To distinguish diphthong and hiatus is necessary an analysis of the near letters around (to left and to right) the vowels involved - this is the context. To demonstrate the concept of context, will be used the follow notation:
  
  - The symbol ‘!’ will indicate a diphthong or a hiatus in the word's end;
  - The symbol ‘&’ will indicate a diphthong or a hiatus in the word's begin;
  - Two letters, like "nç", will indicate a diphthong or a hiatus in the word. The "nç" indicate the letters 'n' and 'ç' enclosures the diphthong or hiatus.

Examples to demonstrate the context concept are below:

- Word: "SA.PU.CAL.A". Context: "a!" (since 'a' is the left near letter of "ia"). Conclusion: The context indicates the two vowels form a hiatus. Every time we have the two vowels "ia", and with the context "a!", this is a hiatus and the syllable must be separated.

- Word: "A.GO.NIA.ÇÃO". Context: "nç" (since 'n' is the left near letter and 'ç' is the right near letter of "ia"). Conclusion: The context indicates the two vowels form a diphthong; thus must be not separated.

- Word: "A.E.RO.DU.TO". Context: "&r" (since 'r' is the right near letter of "ae"). Conclusion: The context indicates the two vowels form a hiatus. Every time we have the two vowels "ae", and with the context "&r", this is a hiatus and the syllable must be separated.

A complete description of the processing of diphthong and hiatus, including a complete list of its contexts, are presented in [Oliveira, 2005].

- "Paroxítonas" words ended with diphthong - In accentuated words in the second syllable from the end to begin of the word ("paroxítona") and terminated with a diphthong, the diphthong is not separated. Here a comment must be done, the grammar and its extensions don't makes a verification of the input text accurate. Incorrect words are processed too and this could produce a wrong output. Correct words are expected.

- Phonetic processing of "AD" prefix - Words with the "ad" prefix and followed by a consonant must have the pronounce as the phoneme "adi" (in portuguese). The vowel 'i' in this case must have pronounced with half time of the normal vowel ('i').

5. The Programs and its Experiments

In [Soares, 2004] were described a program named "fsonora". This program was modified to produce an written output with a delimiter (an dot '.') between the syllables, such as in the word "MA.RA.TO.NA", and provides other parameters to programming, testing and experiment sake. This program doesn't treat accentuation. The program presented in [Soares, 2004] - or better, the grammar implemented by the program - was modified in [Oliveira, 2005] to contemplate the accentuation issues. This program implements the grammar presented in Sect. 3 and was named "nova fsonora" [Oliveira, 2005].

A special program to handle with the follow items was developed and its name is "extensão". The program "extensão" uses the program "nova fsonora" to processing accentuated (and not ones) texts [Oliveira, 2005]:

Hifen, Uruguaiana, V. 30, n° 58, II semestre 2006
The prefixes DES, DIS, TRANS, CIS, SUB, SUPER, HIPER, INTER, AB, and BIS;

The change of 's' between two vowels by an 'z' (only to phonetic pronounce);

The diphthongs, hiatus and pseudo hiatus;

Distinction between diphthongs and hiatus by the context. To this item was used the list of contexts extracted from Michaelis [Michaelis, 2002]. This list of contexts was extracted from a list of 35885 words [Michaelis, 2002] with its syllables separated;

Words with the predominant syllable in the second syllable from the end to begin of the word ("paroxitona");

The prefix "AD". This item was resolved by recording the phoneme "AD" as "ADi", with the sound of 'i' with half time.

The programs used in the experiment realized in this work are listed below:

1. The program "fsonora" is described in [Soares, 2004] and its use is only for experiment sake. This program was modified in a due way to make possible the experiment (the output is written rather spoken);

2. The program "extensão" described above;

3. The program "nova fsonora" described above;

4. A program to log the experiment results of to run the programs "fsonora", "nova fsonora" and "extensão".

The experiment consists in to submit 35885 words [Michaelis, 2002] to the processing of the programs "fsonora", "nova fsonora" and "extensão" (such programs were modified to make its outputs written, and make the experiment possible). Such words were extracted from [Michaelis, 2002]. The log program saves the results of each program according three possible values:

1. If the result (output) is correct, this is a "correct output". To verify the correctness of the result, the processed word is compared with the same word in the list of 35885 words. A "correct output" is obtained if both words (the input and output words) have all the syllables equals;

2. If the result is different of the correct word, but with the same length, this is an "error output";

3. If the result has a length smaller than the correct word, then we have a "defect output". Example: input word: "excluir"; output word: "ecluir". Lack of accent in the output word (if the input word have one) is a "defect output" too. Example: input word: "ábaco", output word: "abaco".

5.1 Results of the experiment

The results of the experiment are showed in the Tables 1, 2 and 3. These results were obtained from the submission of the words list (cited above) to processing of the programs "fsonora", "nova fsonora" and "extensão". The program "nova fsonora" enhances the "fsonora" by the treatment of accentuation and was used in a single mode (used alone) and in conjunction with the program "extensão". The program "extensão" uses the program "nova fsonora" and makes a pos processing to handle with the aspects described above (Sections 4 and 5).

The Table 1 shows that the treatment of accentuation (program "nova fsonora") improves the number of "correct output" in more than 13% (against the program "fsonora"). The program "extensão" handle with some extra aspects (Sections 4 and 5) and use the program "nova
fsonora”. This program improves the number of "correct output" in more than 34% (program "extensão" against "fsonora”).

Table 2 show that the program "extensão" is necessary to reduce the "error output" number. This is due to treatment of the aspects described in Sections 4 and 5.

Table 3 shows that the lack of accentuation treatment (program "fsonora") makes the "defect output" number bigger.

<table>
<thead>
<tr>
<th>Table 1. Percentage of &quot;correct output&quot; of the programs used in the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
</tr>
<tr>
<td>&quot;nova_fsonora&quot;</td>
</tr>
<tr>
<td>&quot;extensão&quot;</td>
</tr>
<tr>
<td>&quot;fsonora&quot;</td>
</tr>
</tbody>
</table>

Table 2. Percentage of "error output" of the programs used in the experiment

<table>
<thead>
<tr>
<th>Program</th>
<th>&quot;Error Output&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;nova_fsonora&quot;</td>
<td>29,90%</td>
</tr>
<tr>
<td>&quot;extensão&quot;</td>
<td>8,91%</td>
</tr>
<tr>
<td>&quot;fsonora&quot;</td>
<td>12,69%</td>
</tr>
</tbody>
</table>

Table 3. Percentage of "defect output" of the programs used in the experiment

<table>
<thead>
<tr>
<th>Program</th>
<th>&quot;Defect Output&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;nova_fsonora&quot;</td>
<td>0%</td>
</tr>
<tr>
<td>&quot;extensão&quot;</td>
<td>0%</td>
</tr>
<tr>
<td>&quot;fsonora&quot;</td>
<td>30,33%</td>
</tr>
</tbody>
</table>

6. Concluding Remarks

This paper has presented a grammar to processing portuguese texts to make a "spoken" version of this text. The grammar was implemented and some aspects of the portuguese language were treated in a special and additional program (program "extensão").

The grammar presented in Sect. 3 can be used to build new applications with voice output in portuguese language. The Sect. 4 presented some aspects of the portuguese language, which are useful to "text-to-speech" process. Section 5 had described the programs developed in this work. The presented results (Sect. 5.1) show the percentage of recognized word (and not ones) in the list of 35885 words that were used in the experiment realized in this work. The experiment was described in Sect. 5. The future efforts are: (a) the grammar improvement to handle with digits and (b) a review of the distinction of diphthong and hiatus by context, to solve a lot of contexts that are ambiguous to differentiate a diphthong and a hiatus, as described in [Oliveira, 2005].

References


