FreP – Automatic extraction of phonological frequency patterns and applications to language acquisition

S. Frota & M. Vigário
Universidade de Lisboa – FL and ONSET-CEL,
FCT-PTDC/ LIN/ 70367/ 2006; http://www.fl.ul.pt/LaboratorioFonetica/FreP
Structure of the presentation

1. Background – the relevance of frequency information in linguistics (a very brief overview)

2. FreP – an electronic tool for extracting frequency information of phonological units
   → general description of the tool
   → demo

3. Application to acquisition studies

4. Conclusion
1. Background: The relevance of frequency information in linguistics

Growing body of research devoted to or taking into account frequency issues
(e.g. Bybee, 2001, 2006; Bybee & Hopper, 2001a; Pierrehumbert, 2002, 2003, 2006…; Moates, Bond & Stockmal, 2002, among many many others…)

Some areas of phonology where frequency has been argued to play a role:

- in phonetic reduction (Jurafsky, Bell & Girand, 2002; Vigário, 2003)
- in the regularization of irregular paradigms (Bybee & Hopper, 2001b)
- in the activation of constraints (Thornton, 1996; Booij, 2005)
- in the emergence and development of grammatical units such as the syllable and the word (to mention but a few recent studies: Beckman & Edwards, 2000; Demuth & Johnson, 2003; Levelt, Schiller & Levelt, 1999/2000; Levelt & van Vijver, 2004; Lleó, 2001a, 2001b, 2004; Ota, 2006; Prieto, 2006; Roark & Demuth, 2000; Stoel-Gammon, 1998; ….)
2. *FreP*

Lack of frequency information of various phonological units and their distribution in Portuguese

*FreP*  
an electronic tool designed to extract frequency information of phonological units at the level of the word and below from texts written in Portuguese
2. FreP – description of the tool

FreP

Very general features

- extracts frequency information of phonological units: prosodic words, clitics, syllables, classes of segments, segments, features, stress

- runs on written texts, following regular orthographic conventions

- optimized for the European variety of Portuguese

- from an ongoing joint project with Fernando Martins (Universidade de Lisboa / Onset-CEL), started in July 2004; grant from FCT, October 2007-2010 (€75.000)

- previous presentations of the tool: Vigário, Martins & Frota, 2005, 2006
  Frota, Vigário & Martins, 2006

- web site: http://www.fl.ul.pt/LaboratorioFonetica/FreP

Very general features

- extracts frequency information of phonological units: prosodic words, clitics, syllables, classes of segments, segments, features, stress

- runs on written texts, following regular orthographic conventions

- optimized for the European variety of Portuguese

- from an ongoing joint project with Fernando Martins (Universidade de Lisboa / Onset-CEL), started in July 2004; grant from FCT, October 2007-2010 (€75.000)

- previous presentations of the tool: Vigário, Martins & Frota, 2005, 2006
  Frota, Vigário & Martins, 2006

- web site: http://www.fl.ul.pt/LaboratorioFonetica/FreP
2. FreP – description of the tool

Very general features (cont.)

- freely available to users for scientific, research, or teaching and learning non-commercial purposes (via e-mail contact)

- presently in Beta mode

- still in progress:
  ✓ improving existing facilities
  ✓ adding new facilities

How it works

a fairly predictable relation between Portuguese orthography and the (lexical) phonology of the language

+ the phonological knowledge of the obligatory segmental phenomena that characterize the language

= identification and segmentation of phonological units
2. FreP – description of the tool

What it does

- automatically extracts (identifies and counts) the following phonological units:
  
  • classes of segments (consonants, vowels, glides; non-consonantal nasality + empty V positions)
  • segments (consonants, glides)
  • POA features for Cs: Labial, Coronal, Dorsal
  • syllables
  • phonological clitics
  • prosodic words (PWs)
  
- provides information on the size of prosodic words & clitics
  - in number of syllables
  - in number of segments

- locates word stress

- provides information on the distribution of stress within words (final, penult & antepenult stress), taking into account the size of words

- provides information on the distribution of POA features within words and taking stress into account
2. *FreP* – description of the tool

*What it does (cont.)*

- provides information on the frequency of different syllable types (CV, V, CVC…)
  - by *position in the word* (initial, internal & final)
  - taking into account the *presence/absence of word stress*
  - taking into account the *position in the word & the presence/absence of stress*

- within the class of phonological clitics, sets *enclitics* and *proclitics* apart

- provides information on orthographic objects
  - nº of orthographic words
  - nº of different words
  - frequency of words
  - nº of characters

- provides a phonetic transcription (SAMPA) – in progress
2. FreP – description of the tool

Phonological knowledge behind the computational system

Segments: e.g. Andrade & Kihm, 1988; Mateus & Andrade, 2000
Syllables: e.g. Andrade & Viana, 1994; Vigário e Falé, 1994; Mateus & Andrade, 2000
Prosodic Words: Vigário, 2003
Clitics: Vigário, 2003

Some relevant options made

- lexical phonology / obligatory phonological processes (e.g. Mateus, 1975; Mateus & Andrade, 2000; Vigário, 2003)
  
  Includes: e.g. glide insertion to break a hiatus; nasalization of vowels and glides; nasal diphthongs formation
  
  Excludes: e.g. optional semivocalization; optional deletions
2. FreP – description of the tool

Some relevant options made (cont.)

- \([k^w]\) and \([g^w]\) = phonologically labialized consonants (instead of \(C+G\))
- glides between vowels are amphisyllabic (e.g. \(\text{areia} \) ‘sand’)
- syllables that do not conform to the general principles of syllable construction display a V-slot position (e.g. \(\text{obter} \) ‘to obtain’ > \(o.bV.ter\))

Some limitations/types of errors

- digits are excluded from the analysis procedures
- acronyms and abbreviations are treated as regular words
- morphosyntactic compounds that have more than one word stress but are written as single orthographic words are computed as forming single prosodic words
2. FreP – description of the tool

Some implemented solutions

• manual intervention on the input file (e.g. deletion of consonants that are not pronounced; full writing of digits, abbreviations and acronyms)

• introduction of rewriting processes for phonological transparency (e.g. *somente* > *so mente* > *só mente*; *muito* > *muinto*)

• exclusion of specific sequences from the application of a particular rule

\[ \text{pezinho} \quad (\text{pe})_\omega (\text{zinho})_\omega \\
\neq \\
\text{vizinho} \quad (\text{vizinho})_\omega \\
\neq \\
\text{vez} \quad (\text{vi})_\omega (\text{zinho})_\omega \]
2. FreP – description of the tool

Evaluation of the tool
- A global evaluation of the tool is still in progress
- The automatic calculations generated by FreP have been checked by hand with regard to:
  - PW identification
  - phon. clitic identification
  - stress location
  - syllable count

based on two corpora (one from adult speech, with 22994 orthographic Ws, and another from child-directed, with 23674 orthographic Ws) – see Vigário, Martins & Frota, 2005; Vigário, Freitas & Frota, 2006

Reliability results:
- 99,935% for PW & clitic identification
- 99,930% for stress location
- 99,709% for syllable count
2. *FreP* – description of the tool

**Evaluation of the tool (cont.)**

For *syllable structure* and *major classes of segments* the automatic calculations have been checked on subsections of three corpora (one from adult speech, with 2796 orthographic Ws, one from child-directed speech corpus, with 4000 orthographic Ws, and one from child speech, with 1003 orthographic Ws (Frota et al, 2005)

**Reliability results:**

- 99.746% for syllable structure
- 99.746% for major classes of segments
2. *FreP* – description of the tool

**Requirements for use**

- takes as input non-formatted, plain text files, based on ASCII ISO 8859-1 (.txt files or similar)
- is written on C++ and uses the Microsoft Visual C++ compiler

**Information display**

*FreP* provides two kinds of user interface:

- windows menu-driven system: frequency counts of the units selected plus examples
- a formatted text file designed to be read by a spreadsheet or database utility (e.g. Excel)
2. *FreP* – description of the tool

**Requirements for use**

- It’s among our plans to make the tool portable to **other platforms**
- It’s part of the FreP Project to adapt the tool to **different varieties of Portuguese** (namely **BP**)
- Look for partners to adapt FreP to **other (Romance) Languages**: Catalan, Spanish

**Sets of functions within FreP**

* Allows to follow the outputs in the formatted text file:

1. **Read and Wipe** the input file
   - ✓ orthographic objects to be ignored deleted; word units divided & listed

   *(column A of the spreadsheet)*
2. *FreP – description of the tool*

**Sets of functions within FreP**

2. **Prepare stress assignment**
   - word tagging for rewriting or exclusion from stress computation; identification of clitics

   *(columns B, C of the spreadsheet)*

   - sozinho > #só# #zino#  
     - ‘alone’
   - vizinho > #vizinho#  
     - ‘neighbour’

3. **Process the phonology of Cs (SAMPA)**

   *(column C of the spreadsheet)*

   - #só# #ziJo#

4. **Stress assignment**

   Applies to units in column C able to feed the process

   *(column D)*

   - #s´o# #z´iJo#
2. *FreP* – description of the tool

**Sets of functions within FreP**

5. *Prepare syllabification*
   ✓ Includes setting V and G apart & identifying the nasal segment that yields nasal Vs and Gs
   *(column E of the spreadsheet)*

   praia > #pr’aGa#
   ‘beach’

   não > #n’ãGN#
   ‘no’

6. *Syllabification*
   ✓ PWs and clitics divided into syllables
   *(column F of the spreadsheet)*

   praia > #pr’aG.Ga#
   ‘beach’

   não > #n’ãG.N#
   ‘no’

   andava > #aN.d’a.va#
   ‘walked’
2. *FreP* – description of the tool

**Sets of functions within FreP**

7. *Major classes of segments*

   ✓ Objects become C, V, G, N; phonological template of the orthographic units in column A
   
   (column G of the spreadsheet)

   ```
   praia > #CC´VG.GV#  
   ‘beach’
   não > #C´VGN#  
   ‘no’
   andava > #VN.C´V.CV#  
   ‘walked’
   ```

8. *Word frequency*

   ✓ List of words by frequency; number of different words
   
   (bottom of the spreadsheet)

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Que</td>
<td>1025</td>
<td>‘that’</td>
</tr>
<tr>
<td>E</td>
<td>742</td>
<td>‘and’</td>
</tr>
<tr>
<td>Não</td>
<td>730</td>
<td>‘no’</td>
</tr>
<tr>
<td>A</td>
<td>691</td>
<td>‘the-FEM’</td>
</tr>
<tr>
<td>E</td>
<td>685</td>
<td>‘is’</td>
</tr>
</tbody>
</table>

   NB: most frequency counts in the spreadsheet
2. **FreP** – Demo

_A note on the data used in this demo_

1. Text composed of CHILDES orthographic lines
2. Because orthographic transcription in CHILDES must allow to recover the target, phonological categories (PW and clitics, major classes of segments, i.e. C, V, G, N) and phonological shapes (nº of syllables per PW/Cl; word stress location; syllable types, also by position in the word and status relative to stress) in child speech and child directed speech can be retrieved by _FreP_.

The actual phonetic form of segments replaced by other segments is not retrieved if it is not transparent from the orthographic form (e.g. "pato" when the child said ['katu] is not retrieved, but “cato” for “pato” is retrieved).
2. **FreP – Demo**

3. For *FreP* to run on the **orthographic output of CHILDES**, the text file has to be wiped off of orthographic characters introduced as special markers (see below; and the *CHAT* manual for a complete list). Most of these characters are automatically removed when *FreP* reads the input text file: namely, punctuation marks and sequences within parentheses and square brackets (presently: +, #, <, >, @x, [ ], ( ) ). The remaining characters have to be removed by the user.

*it is possible to implement the addition of these markers to the automatic wiping off function*

- e.g. xxx/yyyy (unintelligible speech); V@ (proto-determinants)
- :: (pauses between syllables); + (compounds);
- (x)xxxx (segments omitted in production which are part of a word)
2. **FreP** – Demo

Input data: Tomás > TomásCS / TomásCDS

Output spreadsheet: PortugalAnos90
TomásCS / TomásCDS
GuillemCAT(CS)
3. Application to acquisition studies

The role of frequency in Acquisition
3. Application to acquisition studies

- Fundamental research on the phonology of Portuguese
- Teaching/Learning L1 and L2
- The role of frequency in Acquisition
- Language variation: space, time, style, speaker
- Forensic research
- Speech therapy: diagnostic, evaluation
3. Application to acquisition studies

1. What has been done using FreP

   One of the first applications of the tool was in acquisition studies

2. What can be done using the current version of FreP

   Planned in the FreP Project

3. Future developments of the tool

   As planned in the FreP Project (and more…)

Features
- Syllable structure (coda)
- PW shapes: size, prominence
3. Application to acquisition studies

1. What has been done (2004-2007)

1. Vigário, Freitas & Frota (2006) used FreP to extract prosodic word shape frequencies (i.e. frequency of PW with different sizes) both on an adult speech and a child-directed speech corpus.

   It was found that the statistical properties of the input do constrain the shape of early words in Portuguese.

3. Application to acquisition studies

*FreP* was used to obtain overall frequency of

- syllable types
- syllable type distribution by *position within the word* and by *position relative to stress* (e.g. stressed/unstressed)

in different corpora

- adult speech (AS)
- child directed speech (CDS)
- child speech (CS) > ‘*phonetic orthography*’

*It was found that...*

The overall input frequency is **not** a good predictor of the order of emergence of syllable types

- overall frequency predicts the order *CV > V > CVC > CVG/N; ≥*
- the development pattern in child speech displays both *CV and V* at the beginning, and *CVG/N before CVC >>*
3. Application to acquisition studies

The two mismatches (AS, CDS vs. CS) can be explained if the distribution of syllable types by position within the word & by position relative to stress is taken into account.

V and CVG/N are syllable types that appear mainly in prominent positions in the input (i.e. at word-edges and/or in stressed position).
3. Application to acquisition studies

Hypothesis entertained:

- frequency
- structural information (i.e. prosodic prominence)

order of emergence of syllable types in European Portuguese
3. Application to acquisition studies

1. What has been done (2004-2007)

3. Costa, Freitas, Frota, Vigário & Martins (2007) used FreP to investigate a potential correlation between input frequency and developmental patterns for consonantal PoA.

FreP was used to obtain the frequency patterns of PoA in adult speech and child speech data (Labial, Coronal, Dorsal at syllable onset, word initial, medial and final position, and taking stress into account)

It was found that the statistical properties of the input may account for the No *[Dorsal in EP child data: D >> L >> C (D at the PW left-edge)]

4. Pós de Mina (in progress) uses FreP to study the development of the lexicon in child speech
2. What can be done using the current version of FreP

[Runs on (Portuguese) orthography!]

- **Target** [orthographic line of PHON / CHILDES]
  
  ✓ **ALL:** (i) parse the input text into the relevant phonological units > PW, clitic (proclitic, enclitic), syllable, major class of segments (C, V, G), segment type (for C and G), features (PoA for Cs); (ii) assign stress; (iii) compute the frequency of all units (also taking stress and position within the word into account); (iv) provide the frequency of each different orthographic word and the number of different words produced → [not language-specific !] / [≠ CLAN list]
3. Application to acquisition studies

2. What can be done using the current version of FreP

- **Target** [orthographic line of PHON / CHILDES]

  ✓ ALLOWS: (i) phonological description of the target at the word-level and below; (ii) knowledge of the statistical properties of the target: (a) frequency in the **input** and (b) frequency in the **lexicon** (lexicon < list of different words produced); (iii) assessment of the impact of frequency on the acquisition and development of phonology
3. Application to acquisition studies

2. What can be done using the current version of FreP

- **CDS** [CHILDES: more diff. from the target as the coded orthography is closer to actual production]

  ✓ **ALL**: (i) parse the input text into the relevant phonological units > PW, clitic (proclitic, enclitic), syllable, major class of segments (C, V, G), segment type (for C and G), features (PoA for Cs); (ii) assign stress; (iii) compute the frequency of all units (also taking stress and position within the word into account); (iv) provide the frequency of each different orthographic word and the number of different words produced → [not language-specific !]
3. Application to acquisition studies

2. What can be done using the current version of FreP

- CDS [CHILDES: more diff. from the target as the coded orthography is closer to actual production]

  ✓ ALLOWS: (i) phonological description of CDS at the word-level and below; (ii) knowledge of the statistical properties of CDS: (a) frequency in the input (language use) and (b) frequency in the lexicon (lexicon < list of different words produced); (iii) knowledge of how CDS evolves; (iv) assessment of the impact of CDS on the acquisition and development of phonology; (v) comparison between CDS and the target (AS)
3. Application to acquisition studies

2. What can be done using the current version of FreP

- **CS** [phonetic orthography is required, e.g. ‘tato’ for ‘pato’]  
  ✓ With limitations in phonetic detail (at the feature level):  
    (i) parse the input text into the relevant phonological units > PW, clitic (proclitic, enclitic), syllable, major class of segments (C, V, G), segment type (for C and G), features (PoA for Cs); (ii) assign stress; (iii) compute the frequency of all units (also taking stress and position within the word into account); (iv) provide the frequency of each different orthographic word and the number of different words produced → [not language-specific !]
3. Application to acquisition studies

2. What can be done using the current version of FreP

- **CS** [phonetic orthography] is required, e.g. ‘tato’ for ‘pato’

  ✓ ALLOWS: (i) phonological description of CS at the word-level and below; (ii) knowledge of the statistical properties of CS: (a) frequency in language use and (b) frequency in the lexicon (lexicon < list of different words produced); (iii) knowledge of how CS evolves; (iv) comparison between CS and the target (AS); (v) comparison between CS and CDS
3. Application to acquisition studies

3. Future developments of the tool

- As planned in the FreP Project
  - Subclasses of segments
e.g. sonorants, obstruents, plosives, fricatives…
  - All features (PoA, manner, laryngeal …)
  - Full phonetic transcription (SAMPA) from written text
  - Frequency of phonetic segments and their features (also by position within the word/syllable and status relative to stress)
  - Version optimized for the Brazilian variety
3. Application to acquisition studies

3. Future developments of the tool

- As planned in the FreP Project

  ✓ Will allow: (i) a more complete phonological description at the segment and feature level; (ii) the automatic phonetic transcription of the orthography line (PHON), i.e. SAMPA-target / of CHILDES orthographic lines; (iii) comparison between the automatic transcription and the actual production, in case the latter has been transcribed (e.g. between the target and the child production); (iv) comparison between the two major varieties of Portuguese
3. Application to acquisition studies

3. *Future developments of the tool*

- Further developments
  
  i. Orthographic unit to phonological unit correspondence in other (Romance) Languages
     
     e.g. Catalan, Spanish
  
  ii. Explore the FreP concept: adapt the tool to other languages with a grapheme – phoneme relation
3. Application to acquisition studies

3. Future developments of the tool

- Further developments
  - Will allow:
    - (i) assessment of language-specificity/variation in phonology (by applying the set of rules of Language A to the text of Language B), of particular interest in the study of bilingual acquisition;
    - (ii) versions of FreX optimized for different languages (user selects the version he/she wants to use)
4. Conclusion

*Acquisition as an area of application of FreP*

- The tool has already proved to be useful in a number of studies (all of them available at [http://www.fl.ul.pt/LaboratorioFonetica/FreP](http://www.fl.ul.pt/LaboratorioFonetica/FreP))
- some of the planned extensions are acquisition-oriented
- We welcome suggestions on how to make it more useful and usable for research in this area
- We welcome partners interested to test/adapt FreP to other languages

…

*Final remark:* Languages with alphabetic writing systems, and thus a grapheme – phoneme relation, can in principle take advantage of the FreP concept
Work done resorting to FreP – selected references


3.4. Acquisition of prosodic structure: *early words*

- **Grammar**
  Cluster of properties cuing the PW (word bound, edge related, prominence related)
  Presence of early word-based phonology (edge related)
- **Frequency**
  Nice correlation between word shape frequencies in AS and CS (.99)
- **Potential interaction of grammar and frequency**
3.4. Acquisition of prosodic structure: *early words*

- Cross-linguistic implications (though the data is still scarce)
  English, Spanish, Catalan, EP: shapes larger than a binary foot
  Prediction based on **Grammar**: Eng > EP > Sp, Cat
  Prediction based on **Frequency**:
    - Sp (≈30%), EP (27%) > Cat (15%) > Eng (≈5%)
  Reported facts: Sp, EP > Eng, Cat
  **Frequency**, but Eng and Cat?
- If a **Grammar & Frequency** interaction is assumed: a considerably higher frequency in Cat, but much strong grammar cues in Eng.
3.4. Acquisition of prosodic structure: *early syllables*

- **Frequency**
  Overall input frequency is not a good predictor of the order of emergence of syllable types.
  
  **Frequency prediction**
  
  CV > V > CVC > CVG/N
  
  **Order of emergence**
  
  CV, V > CVG/N > CVC

- **Grammar**
  Mismatches explained if prosodic prominence is also taken into account: V and CVG/N appear mainly in prominent positions (word-edges and stressed syl)
3. Application to acquisition studies
(from Vigário et al, 2006)

Syllable types

CV, V, CVC = 73%

CV > V > (C)V(C)> (C)V(N)>(C)VG(N) > CCV
3. Input: word-edges vs. internal position
(from Frota et al, 2005)

ADS - Syllables by position in w

<table>
<thead>
<tr>
<th>Syllable Type</th>
<th>Initial</th>
<th>Final</th>
<th>Internal</th>
<th>Monow</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>V</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CVN</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CVG</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CVGN</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VG</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VN</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CGV</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>GV</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>GVGN</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CVGC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VGN</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>GVC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CVNC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VGC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CVCC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CCV</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CCVN</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CCVGN</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CVGNC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VNC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>VNC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CCVC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CGVC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CGVC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CCVC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CCVC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>GCVC</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Syllable diversity & complexity:
Final>Monow>Initial>Internal
3. Word stress: CV, V
(from Frota et al, 2005)

ADS - The effect of stress
3. Word stress: CVN, CVG

(from Frota et al, 2005)

ADS - The effect of stress
Pós de Mina (in progress): *O léxico de crianças em idade pré-escolar: um estudo de frequências*

**Estudo Preliminar**

<table>
<thead>
<tr>
<th>Idades</th>
<th>N.º Total de palavras únicas</th>
<th>Frequência Lexical</th>
</tr>
</thead>
<tbody>
<tr>
<td>0;11</td>
<td>X=18</td>
<td></td>
</tr>
<tr>
<td>1;2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;4</td>
<td>X=271</td>
<td></td>
</tr>
<tr>
<td>2;8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;11</td>
<td>X=623</td>
<td></td>
</tr>
<tr>
<td>3;2</td>
<td>X=886</td>
<td></td>
</tr>
<tr>
<td>3;4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4;2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4;4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4;7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4;10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Desempenho Lexical**