

# **LumaLiDaOn**

Frota, S., M. Vigário & R. Jordão (2008) – Version 1

Frota, S., M. Vigário & R. Jordão (2012) – Version 2

Laboratório de Fonética, CLUL/FLUL

## **Transcription Criteria**

1. Child Utterance
2. 'Phonetic' Orthography of child utterance
3. Target Utterance
4. Target Orthography
5. Meaning
6. Prosodic Phrasing
7. Observations on other fields of the database

# 1. Child Utterance

## 1.1. On the nature of the phonetic transcription in LumaLiDa

Phonetic transcriptions in LumaLiDa have two main properties: they are online transcriptions and were made in the context of utterance production.

Child utterances were phonetically transcribed online by one of the parents, as they were produced. The initial transcription was made in IPA. When both parents were present at the moment of production, the non-transcriber has checked the transcription. Any disagreement was signaled and discussed afterwards. Uncertainty about the nature of a segment or the presence of a segment (by one of the transcribers or both) is signaled with a special code (see Special codes below).

On the basis of their daily experience, transcribers find their phonetic transcription reliable in what concerns the presence/absence and nature of segments, of their syllabic status (i.e. vowels and glides, particularly in case of diphthongs), and of word stress. It should be noted that it was not among the goals of the transcription to provide fine phonetic detail, such as secondary articulation features (like gradient palatalization or rounding of consonants, and gradient vowel raising or lowering). In this respect, a decision was always made based on the closest segment in the target system.

The phonetic transcription of child utterances was converted to SAMPA. All phonetic transcription fields in the database use the Portuguese SAMPA ([www.phon.ucl.ac.uk/home/sampa/portug.htm](http://www.phon.ucl.ac.uk/home/sampa/portug.htm)). Any additional symbols used are indicated below (see Special Codes).

## 1.2. General criteria

- (a) Word stress is marked immediately before the vowel, even in rising diphthongs.
- (b) Words (corresponding to orthographic words) are separated by a blank space, except in the case of pronominal enclitics.
- (c) All prosodic words are marked with a primary word stress.
- (d) The segments usually taken to be labialized in the target – [K\_w] and [g\_w] – were always transcribed as consonant plus glide (Cw) in child speech: e.g. [kw"atu], target [k\_w"artu].
- (e) The child uttered a velar-pharyngeal nasal, transcribed as [N\*]: e.g. [s"ON\*], target [s"Ol\_G].
- (f) The presence of an empty position that has its own timing slot is indicated with a `+' : e.g. [ko~+g"iw], target [ko~s@g"iw].
- (g) When the same sequence is uttered several times and the number of times it was uttered is indicated by the transcriber, the entry is repeated the exact number of times the sequence was produced; when the same sequence was uttered several times by the child but the transcriber has only indicated this fact without exactly stating the number of repetitions, only one entry is made and a comment is inserted in the Comments field.

## 1.3. Special codes

- (a) Bold-yellow marks uncertainty relative to the nature of the segment(s) produced (often more than one possibility is considered as indicated in the Comments field).

- (b) Curved parenthesis marked uncertainty relative to the production of that particular segment.
- (c) § signals uncertainty relative to the presence of a segment at that point in the utterance.
- (d) underlying highlights segments that were clearly produced as such, in a target-like manner, and that usually were not produced as such before or in other instances of the same target sequence.
- (e) Short and medium length pauses are indicated with '...' and long pauses with '//'.

#### 1.4. Additional observations

The phonetic transcription field may be empty in entries showing only the orthographic fields and the target utterance transcription. This situation arises from cases where only the orthography was registered in the diary. These entries were kept in the database as they may be a valuable source of information for morphological, lexical, syntactic and semantic purposes.

## 2. 'Phonetic' Orthography of child utterance

### 2.1. Why a 'phonetic' orthography

This field was introduced to enable the tool FreP ([www.fl.ul.pt/laboratoriofonetica/](http://www.fl.ul.pt/laboratoriofonetica/) FreP) to run on child data, so it is of special interest both to FreP users and to researchers that want to take advantage of the facilities provided by FreP to get the frequency of phonological objects and patterns out of child speech data.

### 2.2. General criteria

- (a) The orthography includes only the segments that are present in the phonetic transcription, i.e. the number of segments in both transcriptions matches.
- (b) Coalescence is respected unless it is independently known that the child is able to produce all the elements involved in the process.
- (c) Orthographic marks of stress are inserted to ensure that the location of stress in speech is computed correctly. Thus these marks signal stress position and not vowel quality.
- (d) A single word may not have more than one orthography stress mark (acute stress).
- (e) Non-full vowels in the phonetic transcription are included in the orthography.
- (f) Deleted vowels are excluded from the orthography.
- (g) The insertion of the [ə] sound is transcribed as 'e' in the orthography: e.g. [f'ʊ@6] > fóea.
- (h) The nasal [ɨ] is not indicated in the orthography: e.g. [dij~"u] > diiu.
- (i) Target names are written with an initial capital letter.

## 3. Target Utterance

### 3.1. On the nature of the phonetic transcription of the target

The phonetic transcription of the target is generally conservative in the sense that it closely follows the orthography when more than one option is available in the language.

By and large, the transcription reflects the lexical phonology. External sandhi phenomena are included only when obligatory or very frequent.

### 3.2. General criteria

- (a) On the lexicon: the target lexicon is indicated and not specific forms used by the child (e.g. child 'nana' > amarelo; child 'pepé' > sapato).
- (b) If clear from the child utterance and context, grammatical information like number and gender are included.
- (c) Clitic words are also included.
- (d) The word order of the language is followed.
- (e) Word initial vowels are transcribed following the orthography (e.g. [el@f'6~t@] 'elefante', [obrig"ad6] 'obrigada').
- (f) Other cases with possible alternative productions are also transcribed following the orthography (e.g. 'piscina' with two [i] vowels and 'crescido' with one [@] and one [i] vowel).
- (g) The segments usually taken to be labialized in the target – [K\_w] and [g\_w] – are transcribed as such.
- (h) Voiced plosives are not transcribed as non-strident fricatives in between vowel position.
- (i) The 'ing' sound is transcribed as a velar nasal ([N]).
- (j) Ambisyllabicity is not marked.
- (k) All prosodic words are marked with a primary word stress.
- (l) The homophonous forms of the indefinite article and the numeral ('um', 'uma') are unstressed in the case of the article and stressed in the case of the quantifier and in case of doubt between one and the other.

## 4. Target Orthography

The target orthography is simply the orthographic counterpart of the target phonetic transcription. Proper names are given with an initial capital letter.

## 5. Meaning

This field is intended to make clear what the exact meaning of the child utterance is. It thus may correspond to the target orthography or include additional elements explaining the intended meaning of the utterance.

## 6. Prosodic Phrasing

Labelling for prosodic phrasing followed the general properties of the prosody of Portuguese (Frota 2000, 2014; Vigário 2003), together with information on the specific features of Luma's speech as described in Frota & Vigário (2008), Frota & Matos (2009). As a rule, we opted for minimal phrasing, and thus the prosodic phrasing deviates from adult prosodic structure in three ways:

- (1) optional phrasing effects were not considered (e.g., weight effects and compound prosodic domains);
- (2) a PW was usually mapped into PhP, following the tendency described in Frota &

Vigário (2008) and Frota & Matos (2009);

(3) prosodic levels between PW and PhP were not considered, beyond the assumption that the unstressed syllables of clitics adjoin to the adjacent PW.

## 7. Observations on other fields of the database

The Context field includes information on the context of production (such as previous and/or following discourse, description of the scenario, or description of actions).

The Comment field comprises all sorts of comments relevant to the nature of the phonetic transcription and/or the meaning of the utterance produced.

The SemPragObs (Semantic & Pragmatic Observations) field only contains, for the moment, information on the sentence type of child utterances, as indicated in the diary by the transcriber(s).