

## LUDViC Itags

### 1. Introduction

The linguistic tags (Itags) used so far consist in an annotation level about phonetic/phonological properties which distinguish variants of the same lexical item and that in this corpus are reflected in the orthographic representation (transcription).<sup>1</sup> Some of these Itags have a common logic, but others have one of their own; this depends on the property to be signalled, as is described below. Before that, two important observations are needed:

1. No analysis is provided here as to the type of change that originated these current variants; each Itag only signals a distinction among them, thus enabling specific searches which can be useful for dedicated studies and analyses in the future.
2. Each lexical item can receive more than one Itag. Some of these are independent, but in other cases complex phenomena may be captured when we search for certain combinations: e.g. the combinations ‘a-o, C-fin’ or ‘a-e, C-fin’ (both like this, separated with a comma), under the ‘Itag’/ ‘matches’ button, bring results that may be interesting to the study of metaphony.

### 2. List, meaning, and examples

| name       | meaning  | example                                 |
|------------|--|---|
| a-e        | an <i>e</i> occurs in a tonic syllable, where other variants have <i>a</i>                         | <i>seb (sabi)</i>                       |
| a-o        | an <i>o</i> occurs in a tonic syllable, where other variants have <i>a</i>                         | <i>korr (karru)</i>                     |
| e-i        | an <i>i</i> occurs in an unstressed syllable, where other variants have <i>e</i>                   | <i>bibida (bebida), ninhun (nenhun)</i> |
| dj-j-i     | dj-j-i alternation   | <i>djá / já / íá</i>                    |
| tx-x       | tx-x alternation   | <i>katxupa / kaxupa</i>                 |
| v-b        | a <i>b</i> occurs where other variants have <i>v</i>   | <i>baka (vaka)</i>                      |
| C-fin      | a consonant occurs in a final position, where other variants have a vowel following it             | <i>amig (amigu), net (netu)</i>         |
| C-in       | a consonant occurs in an initial position, where other variants have a vowel preceding it          | <i>gora (agora), té (até)</i>           |
| diph       | a diphthong occurs where other variants have vowels (even if signalled with <i>n</i> for nasality) | <i>entāu (enton), ideia (idea)</i>      |
| int-sC     | there is a sC sequence in an internal position   | <i>adolexsent, krexsiment</i>           |
| fin-stress | stress falls on the last syllable  | <i>pensá (pensa), purkê (purki)</i>     |

<sup>1</sup> None of them is therefore defined through the comparison of a variant with the respective lexeme – the lexeme is just an underlying tool that the system needs in order to recognise and connect all the variants of the same lexical item; it thus relates to the core identity of each lexical item, not to any phonological features. And so the lexeme may be used to search in the corpus for all the variants of the same lexical item, but the relevant phonological comparison signalled by the Itags is between the variants themselves. Note that some of these other variants may be still unavailable in the corpus, but we know that they exist in the language. For inquiries on these absent variants, please send an email to: LUDViC2020@gmail.com.

## 3. Description

### 3.1. *a-e*

This ltag is used to signal some cases, mostly typical of the northern varieties, in which an [e] occurs in a tonic syllable while other variants have an [a] in that same position.

### 3.2. *a-o*

This ltag is used to signal some cases, mostly typical of the northern varieties, in which an [o] occurs in a tonic syllable while other variants have an [a] in that same position.

### 3.3. *e-i*

This ltag is used to signal some cases, mostly typical of the southern varieties, in which an [i] occurs in an unstressed syllable while other variants have [e] in that same position.

### 3.4. *dj-j-i*

This ltag is used to signal the alternation between these phonemes (*i* is a semi-vowel, [j]) at the onset of some syllables.

### 3.5. *tx-x*

This ltag is used to signal the alternation between these phonemes at some syllables' onset.

### 3.6. *v-b*

This ltag is used to signal some cases where a *b* occurs as a syllable onset while other variants have a *v* in that same position.

### 3.7. *C-fin*

This ltag is used to signal the ending of a word in a consonant (mostly typical in the northern varieties) while other variants have this same consonant followed by a vowel (generally [i] or [u]).

### 3.8. *C-in*

This ltag is used to signal the starting of a word in a consonant (mostly typical in the southern varieties) while other variants have this same consonant preceded by a vowel.

### 3.9. *diph*

This ltag is used to signal the existence of diphthongs in contexts where other variants have vowels. These diphthongs may be either oral or nasal, but they are certainly falling (so what is missing in the vowel-only alternate is the second element of the diphthong; this does not mean, however, that the pronounced vowel is the same that exists in the diphthong version – sometimes it is not, as we see by the first example in the table on page one).

### 3.10. *int-sC*

This ltag is used to signal variants with the rare combination sibilant + consonant in a word-internal position.

Note that *sC* clusters are very common in the language in a word-initial position – *skola*, *special*, *stá* – in which the sibilant, generally followed by a stop consonant, is optionally pronounced as [s] or [ʃ]. In this case they are always written as /s/, though, as are all the sibilants in a final position, independently of their contextual pronunciation.

This ltag applies to the much less common internal position, where the sibilant is always pronounced as [ʃ] and the following consonant itself is another sibilant, [s] – they are therefore transcribed as *xs*, as in *naxsid*, with the alternative variant rather containing a sequence of regular syllables (consonant + vowel): *nasidu*.

### 3.11. *fin-stress*

This name refers to the stress of a variant falling on the syllable containing the final mora.

In fact, the language “displays a general weight-sensitive stress rule that assigns primary stress to the syllable containing the penultimate mora”, with nasal vowels in a final position behaving as long vowels and thus counting as two moras, and also glides (semivowels) in diphthongs and coda segments counting as moraic (see Pratas 2021:50-51 for a refinement – and an adaptation to all varieties of the language – of this rule, which was first proposed in Pratas & Salanova 2005 for the variety of Santiago).

Some words, however, violate this general rule, with their stress falling on the syllable that contains either the antepenultimate mora or rather the last/final mora (all these are transcribed with a graphic accent on their stressed syllable).

This ltag, as the name indicates, signals only the words which belong to the latter group – i.e. those which end in an oral vowel and yet are stressed on this final syllable – and have variants that obey the general stress rule in the language. This is typically the case of verbs in the northern varieties, but also includes words like *purkê*, a conjunction.

Note that, in the case of the verbs, the ltag applies only to their form with no object clitic or temporal postverbal morphology (e.g. in participles) – the presence of these morphemes changes the composition of the last syllable, and so the previous violation of the language stress rule no longer exists (i.e. the word no longer ends in an oral vowel, and so even if the stress falls on the last syllable, it is here in compliance with the referred general rule).