Syllables and Stress

This page describes how to generate stress-marked syllable annotations after forced alignment, which can be achieved if your original poronunciation dictionary includes syllable/stress marking (e.g. CELEX and Unisyn do, but the CMU Pronouncing Dictionary doesn't).

What LaBB-CAT does is:

1. Take the phonemic transcription of each word from the *segment* layer (e.g. the one chosen by HTK during forced alignment), which has no syllable or stress marks.
2. Look it up in CELEX.
3. If it's found, it takes the syllable/stress-marked version of the phonemic transcription
4. It splits the transcription into syllables, and creates an annotation for each syllable that spans the corresponding phones on the *segment* layer.

Once you have such annotations, it's much easier to identify, for example, specific vowels but in stressed syllables only.

Prerequisites

- A layer that tags each word token with its phonemic transcription according to CELEX or a Unisyn lexicon.
- Force-alignment using the above phonemic transcription layer, so you have a segments layer filled in with aligned phones.

NB: If phone alignments are produced by a different dictionary, then the resulting phonemic transcriptions will not match the dictionary and so syllables cannot be reconstructed. e.g. if you use the MFA built-in English dictionary, with MFA's pretrained acoustic models, reconstructing syllables from CELEX transcriptions will fail.

Steps

1. In LaBB-CAT, click the *word layers* menu option
2. At the top of the list of word layers, fill in the blank header form in order to add a new layer. Important points are:
   - *Type = Phonological*
   - *Alignment = Intervals*

- *Manager = CELEX English*

description	Syllables	**layer type**	Phonological ⌄	**alignment** Time Intervals ⌄
short description	syllables	**layer manager**	CELEX English ⌄	**generate** always ⌄
notes	Aligned, stress marked syllables.			**project** ⌄

◇ New

3. Press *New.* You will bee taken to the layer configuration page.
4. On the left, select the *Syllables from Phonology* option.
   This will automatically set *Source Layer* to *segments* and *Delimiters* to – (hyphen)

Source layer: segments

Language: en.*

Delimiters: -

Pronounce Event Override: ☐

Generates Segments: ☐

Lookup Definition:

Presets   **CELEX SQL Query**

                                            SELECT DISTINCT BINARY PhonStrs
                                             FROM cxen_wordformphonologyprc
○ Phonology   WHERE BINARY REPLACE(REPLACE(F
○ Syllable Count   ORDER BY LOCATE('\'', PhonStrs
○ Morphology
○ Syntax
○ Lemma
○ Frequency (wordform)
○ Frequency (lemma)
◉ Syllables from Phonology
☐ First match only

Test word: testing

[Save]

5. Press *Save*
6. Pres *Regenerate*
   You will see a progress bar as LaBB-CAT processes all the transcripts in the database.
   This processing is done in the background, you don't need to wait until it's finished

before visiting other pages.

Once the layer annotations are generated, each word that has been previously force-aligned will now have one or more syllable annotations, marking out the phones that belong to each syllable.

The label of each syllable annotation is the stress-marked phonemic transcription of the syllable, so for syllables with primary stress, the first character in the label is ' (an apostrophe), and those with secondary stress have " (double-quote) as the first character.
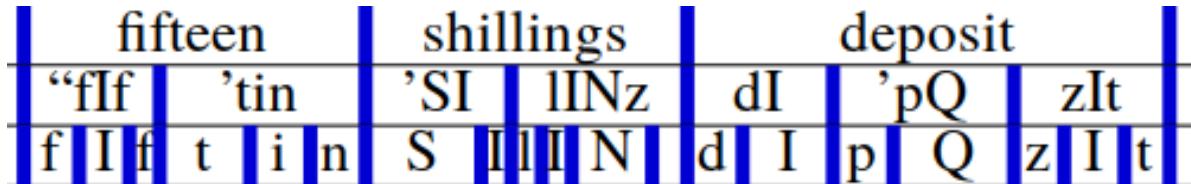

Figure 1: Stress-marked syllables as seen in Praat

Now, if you want to, for example, identify all kit vowels in syllables with primary stress only, you can do a search:

- for *segment*s labelled as the kit vowel
- whose *syllable* starts with the primary stress mark '
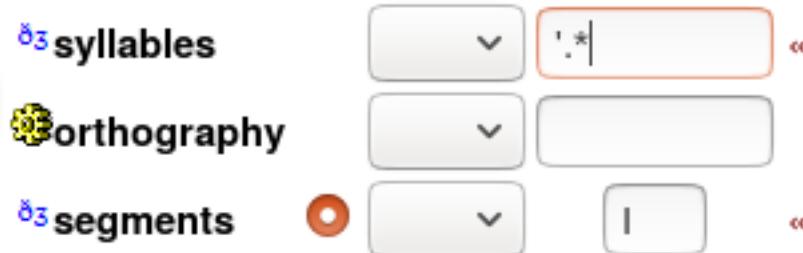
Figure 2: Search matrix for KIT segments whose syllable starts with the primary stress marker