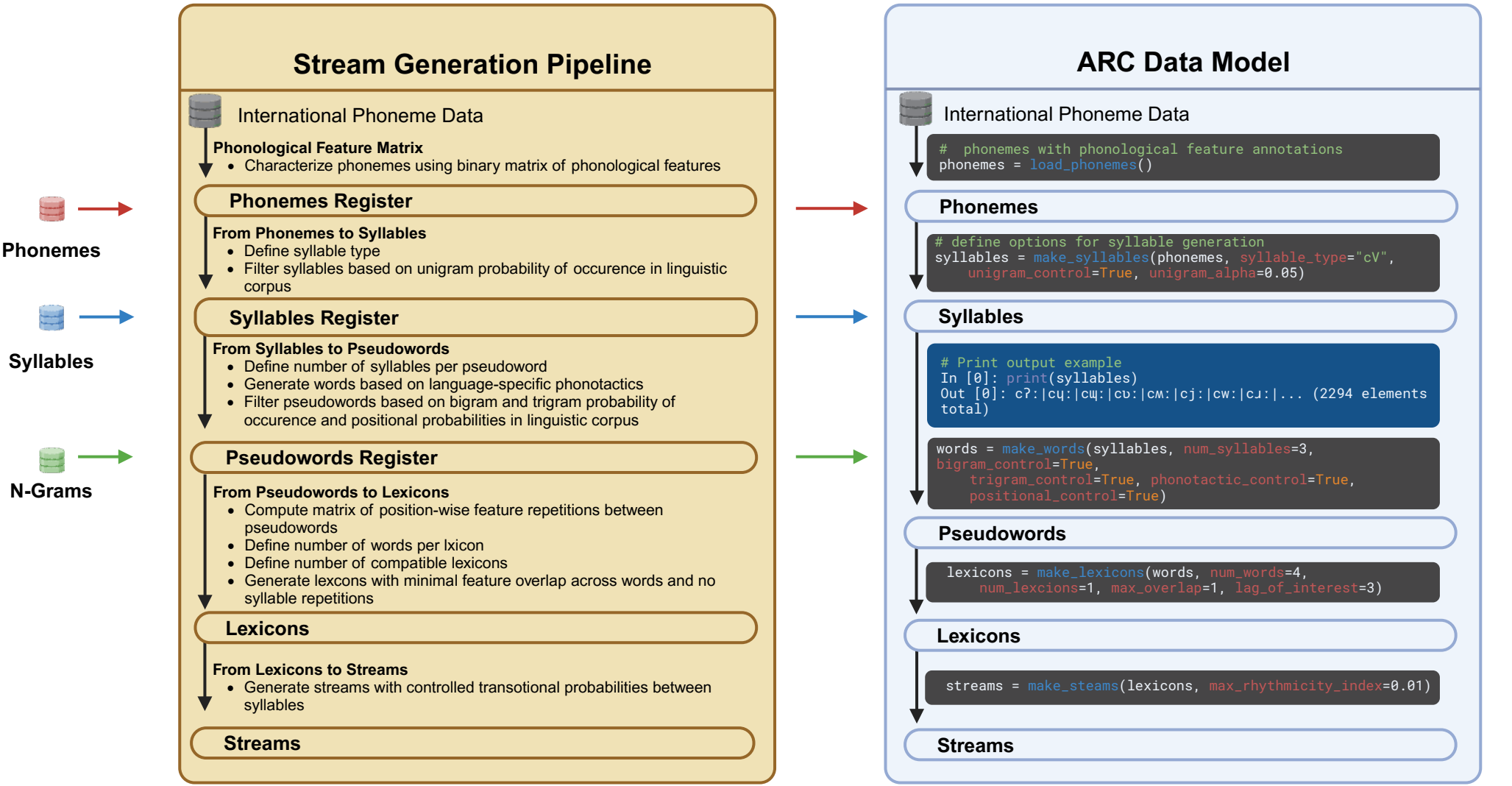
Phonemes
Syllables
N-Grams
Stream Generation Pipeline
International Phoneme Data
Phonological Feature Matrix
• Characterize phonemes using binary matrix of phonological features
Phonemes Register
From Phonemes to Syllables
• Define syllable type
• Filter syllables based on unigram probability of occurence in linguistic corpus
Syllables Register
From Syllables to Pseudowords
• Define number of syllables per pseudoword
• Generate words based on language-specific phonotactics
• Filter pseudowords based on bigram and trigram probability of occurence and positional probabilities in linguistic corpus
Pseudowords Register
From Pseudowords to Lexicons
• Compute matrix of position-wise feature repetitions between pseudowords
• Define number of words per lxicon
• Define number of compatible lexicons
• Generate lexcons with minimal feature overlap across words and no syllable repetitions
Lexicons
From Lexicons to Streams
• Generate streams with controlled transotional probabilities between syllables
Streams
ARC Data Model
International Phoneme Data
# phonemes with phonological feature annotations
phonemes = load_phonemes()
Phonemes
# define options for syllable generation
syllables = make_syllables(phonemes, syllable_type="cV",
unigram_control=True, unigram_alpha=0.05)
Syllables
Print output example
In [0]: print(syllables)
Out [0]: cʔ:|cɥ:|cɰ:|cʊ:|cʍ:|cj:|cw:|cɹ:|... (2294 elements total)
words = make_words(syllables, num_syllables=3,
bigram_control=True,
trigram_control=True, phonotactic_control=True,
positional_control=True)
Pseudowords
lexicons = make_lexicons(words, num_words=4,
num_lexcions=1, max_overlap=1, lag_of_interest=3)
Lexicons
streams = make_steams(lexicons, max_rhythmicity_index=0.01)
Streams