Phonotactic gaps
Accidental gaps: don’t violate any phonotactic restrictions. Systematic gaps: violate some phonotactic constraints(s).

Traditional grammatical approaches presume a categorical distinction between systematic and accidental gaps:
- all systematic gaps are equally ill-formed;
- all accidental gaps are equally well-formed.

This predicts categorical wellformedness judgments.

But: not all unattested words are judged identically!
- Acceptability of unattested words is gradient
- Acceptability reflected in statistical properties of the lexicon (n-gram probabilities, neighborhood density, etc.)

Previous studies focused on accidental gap acceptability, perhaps assuming systematic gaps are equally ill-formed [1] [2] [4] [6]

Research questions:
1. How do Cantonese speakers judge the wellformedness of systematic gaps?
2. Do the judgments correlate with lexical statistics?

Cantonese
(C)G(V)(C) syllable structure
19 onsets: /p t k m n/ 8 monophthongs: /a i u: au ei ru oy ai iu ou/
11 diphthongs: /ai au ei au oy ai iu ou/
6 tones /5 25 33 21 23 22/

Typology of systematic gaps
- Labial dissimilation gaps
  - No labial onsets and labial codas (*pap, *tupap)
  - No labial codas and rounded vowels (*yym, *ymo)
  - No labial onsets and front round vowels (*ymo, *ymo-)
- Onset-tone gaps
  - No aspirated onsets with 22 tone (*pahu22, *thu22)
  - No unaspirated onsets with 21/23 tones (*pau23, *tau21)
- Coronial gaps
  - No coronal onsets and codas with /t s l/ (*tax, *natx)
  - No coronal onsets with /t s l/ (*taxp, *nax)

Experimental corpus
432 items conforming to a CV(C) template, derived from all possible combination of
- eight onset phonemes /p t k m n l t\theta n/
- three vowel phonemes /a i u/
- an optional /m/ /oda/ six tones /5 25 33 21 23 22/
Produces 162 attested syllables and 270 nonwords:
- 61 fill labial dissimilation gaps
- 36 fill onset-tone gaps
- 42 fill coronial gaps
- 27 syllables filled two types simultaneously, 1 all three
Remaining 103 nonwords classified as accidental gaps.

Procedure
Ten Cantonese native speakers were presented with a randomized series of items from the corpus & given two tasks per stimulus:
- Lexical decision: “Is this a word of Cantonese?” (y/n)
- Wordlikeness rating: “How good a word of Cantonese is this?” (1-7, 1 = worst, 7 = best)

Lexical statistics
Phonotactic probability (PP) operationalized as average bigram log probability (1):

\[ P(W) = \frac{\sum_{w \in W} \log p(w|w_{-1})}{|W|} \]

Neighborhood density (ND) operationalized as Levenshtein edit distance between strings

ND(w) = number of syllables in the Chinese Character Database [3] which could be formed by changing, adding, or deleting a single segment (or tone) of w; weighted by token frequency in the Hong Kong Cantonese Adult Language Corpus (HKCAC: [5])

References