A system for unified corpus analysis

applied to duration compression effects across 12 languages

Michael McAuliffe,
Morgan Sonderegger, Michael Wagner

McGill University

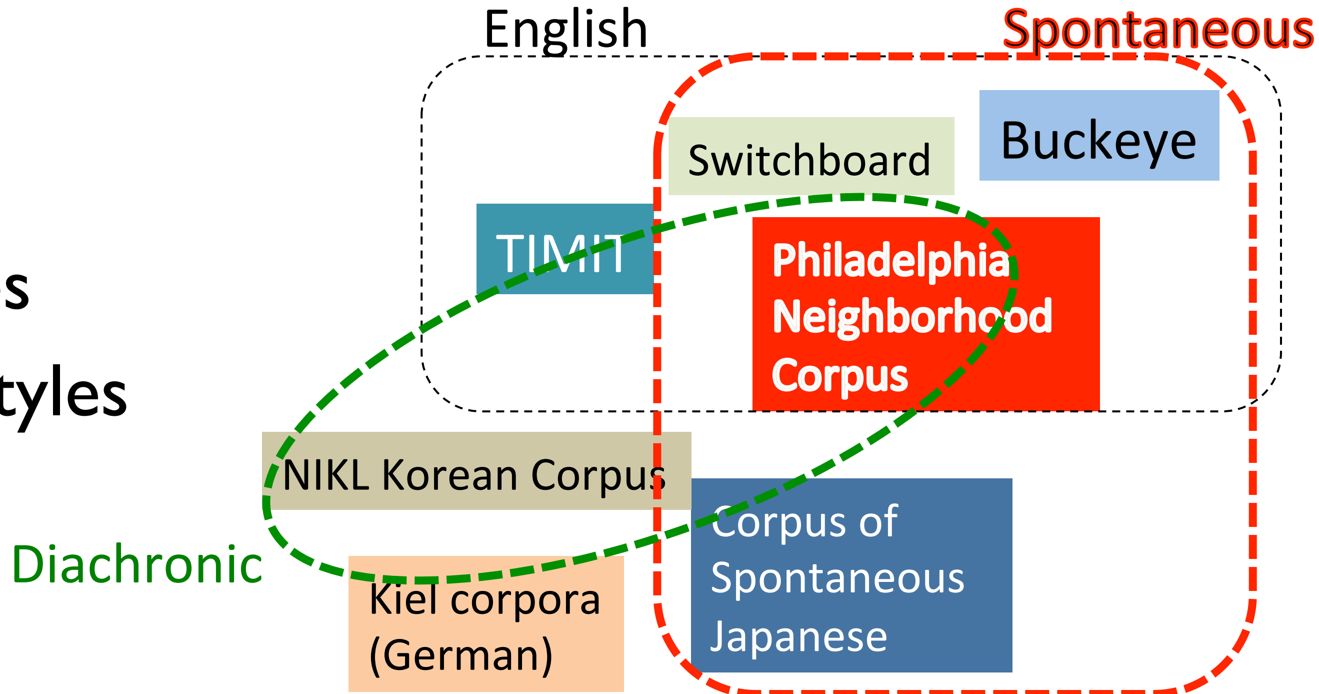
MOLT 2016

Introduction

- Huge amount of annotated speech data exists
  - Corpora
  - Academic labs
  - Web

At least orthography + audio

- Across
  - Languages
  - Speech styles
  - Time



English

Spontaneous

Diachronic

Switchboard

Buckeye

TIMIT

Philadelphia Neighborhood Corpus

NIKL Korean Corpus

Corpus of Spontaneous Japanese

Kiel corpora (German)

Introduction

- Great potential for phonology/phonetics
  - Bigger haystacks, same-sized needle…
  - ... need a bigger magnet
- Requires software for unified corpus analysis
  - Integrating speech datasets
  - Querying across them
- Today: Speech Corpus Tools
  - Case study: duration compression effects in 12 languages
  - Yesterday: application to Buckeye (Kilbourn-Ceron et al.)

Why is using corpora hard?

- Speech datasets:
  - Large
  - Complex
  - Diverse formats

- Access to many speech datasets
  - Costly or ethically restricted

Most sociolinguistic, fieldwork, laboratory data

Switchboard: $3000+

- Result: requires lots of specialized code, $$, effort

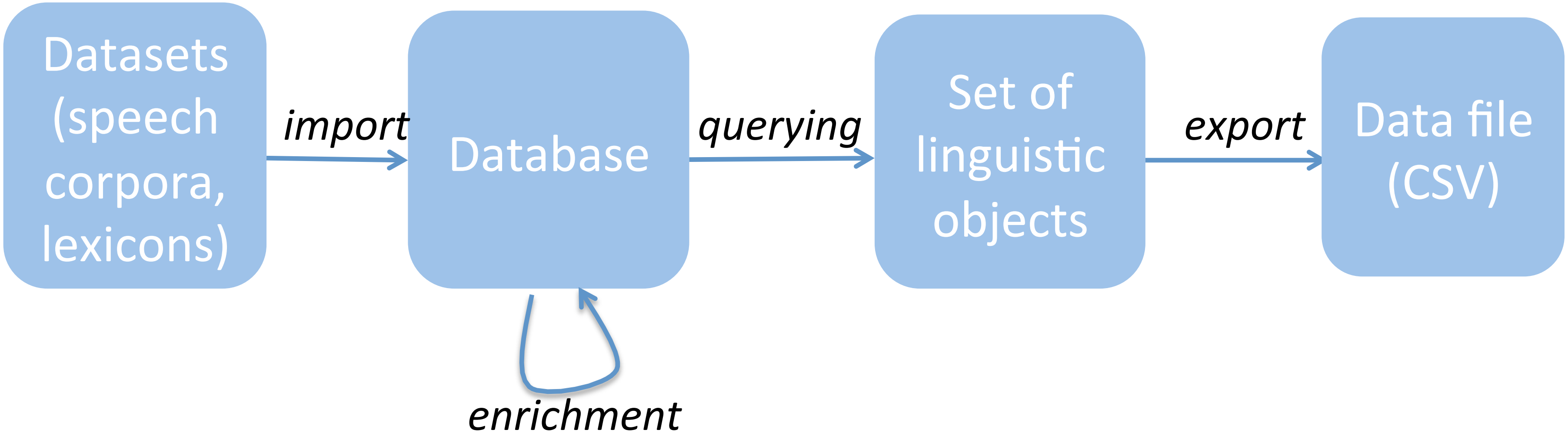
Related work

- Phon (Rose et al., 2007)
  - Construction + querying of individual speech corpora
- EMU (Cassidy & Harrington, 2001)
  - Annotation, integration, querying
- Annotation graphs, ATLAS (Bird & Liberman, 2001; Bird et al., 2000)
  - Formal model for linguistic annotations
  - Linear signals (e.g. speech)

SCT: Goals

- Scalable
- Require minimal technical skill from user
- Abstraction away from dataset format
- Querying dataset without access to raw data

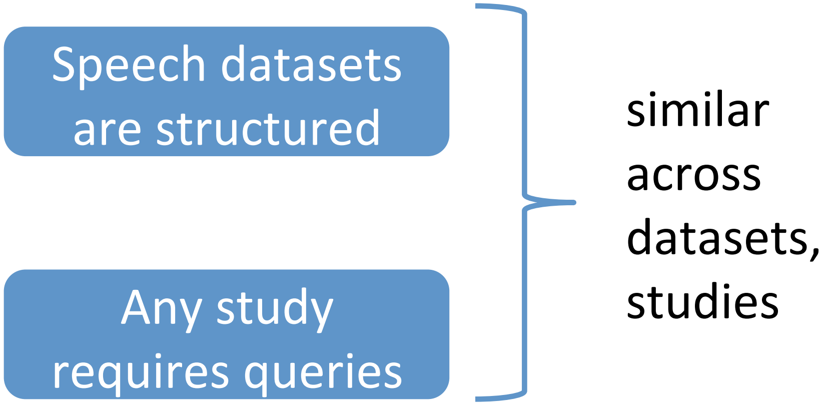
SCT: structure


Datasets (speech corpora, lexicons)
import
Database
enrichment
querying
Set of linguistic objects
export
Data file (CSV)

- Implementation
  - Python module
  - Graphical interface (release: LabPhon 2016)

SCT: Databases

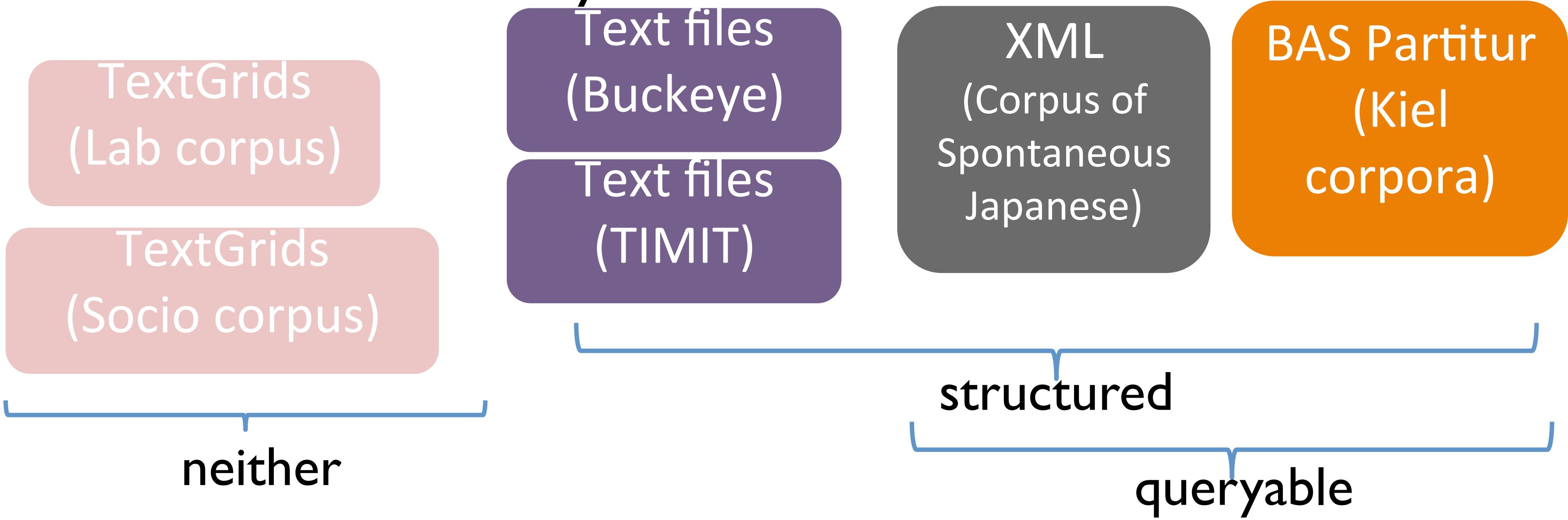
- Why databases?
  - For structured data: organization || structure
  - Queryable
  - Standarized way



Speech datasets are structured

Any study requires queries

similar across datasets, studies



TextGrids (Lab corpus)

TextGrids (Socio corpus)

Text files (Buckeye)

Text files (TIMIT)

XML (Corpus of Spontaneous Japanese)

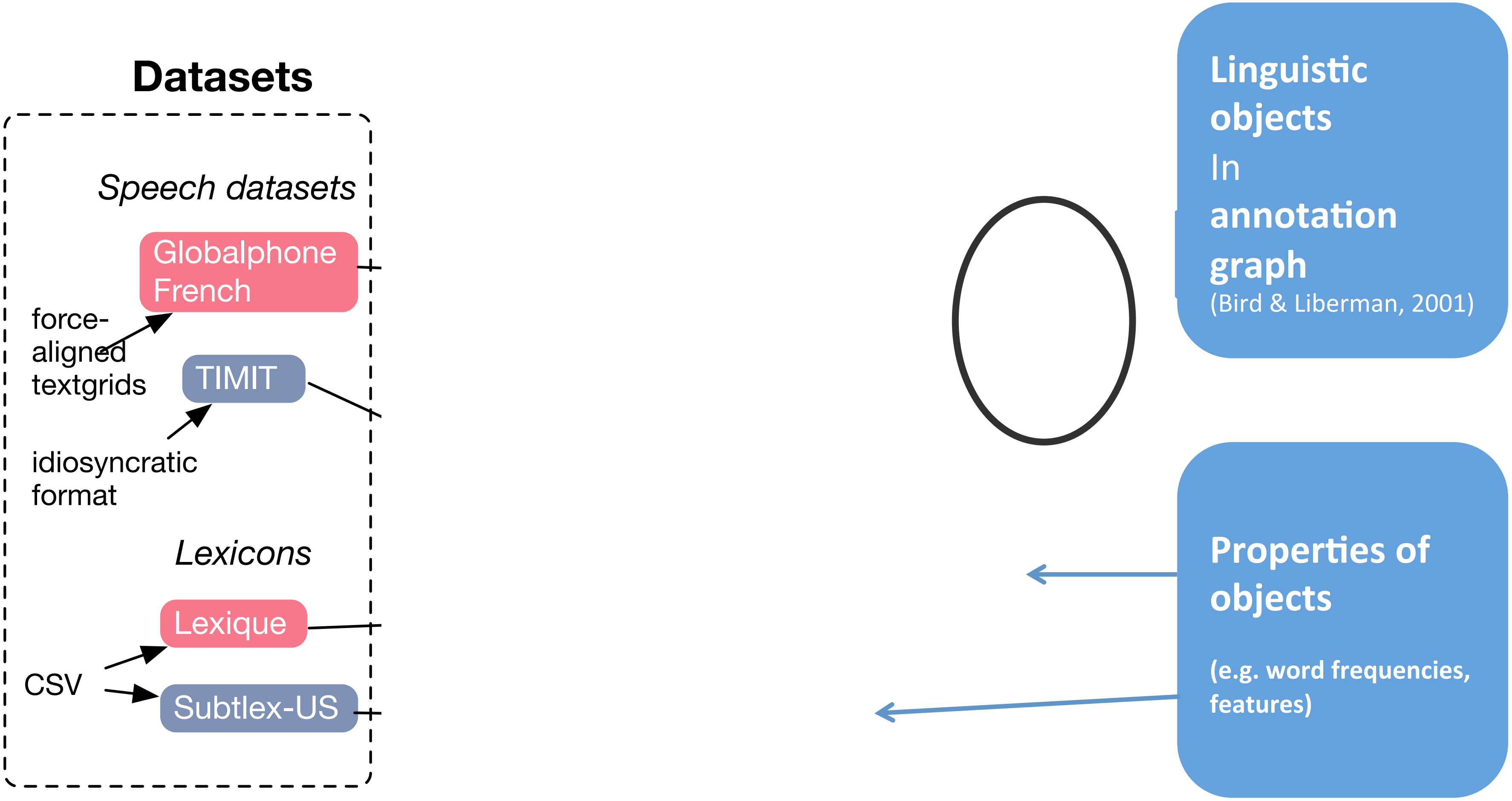
BAS Partitur (Kiel corpora)

structured

queryable

neither

SCT: Import



Datasets

Speech datasets

Globalphone French

force-aligned textgrids

TIMIT

idiosyncratic format

Lexicons

Lexique

CSV

Subtlex-US

Linguistic objects

In **annotation graph**

(Bird & Liberman, 2001)

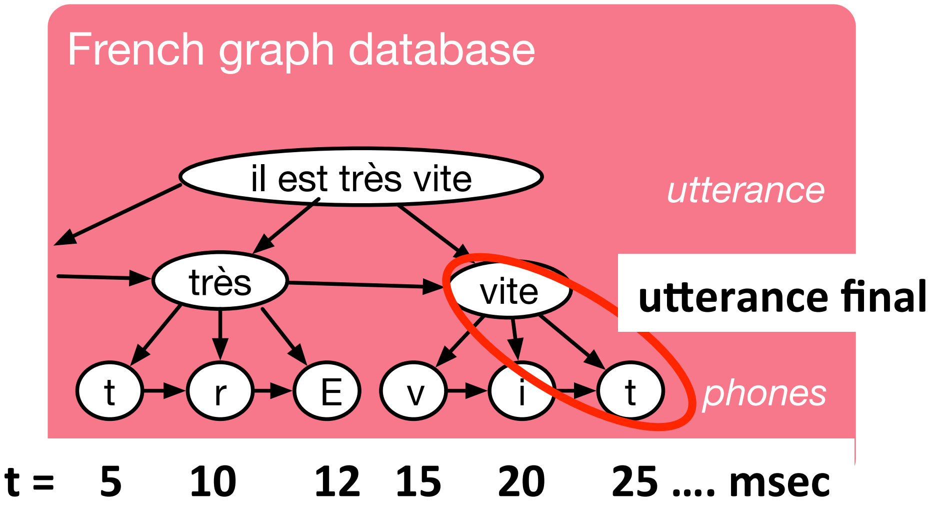
Properties of objects

(e.g. word frequencies, features)

- Speech, text datasets → queryable databases

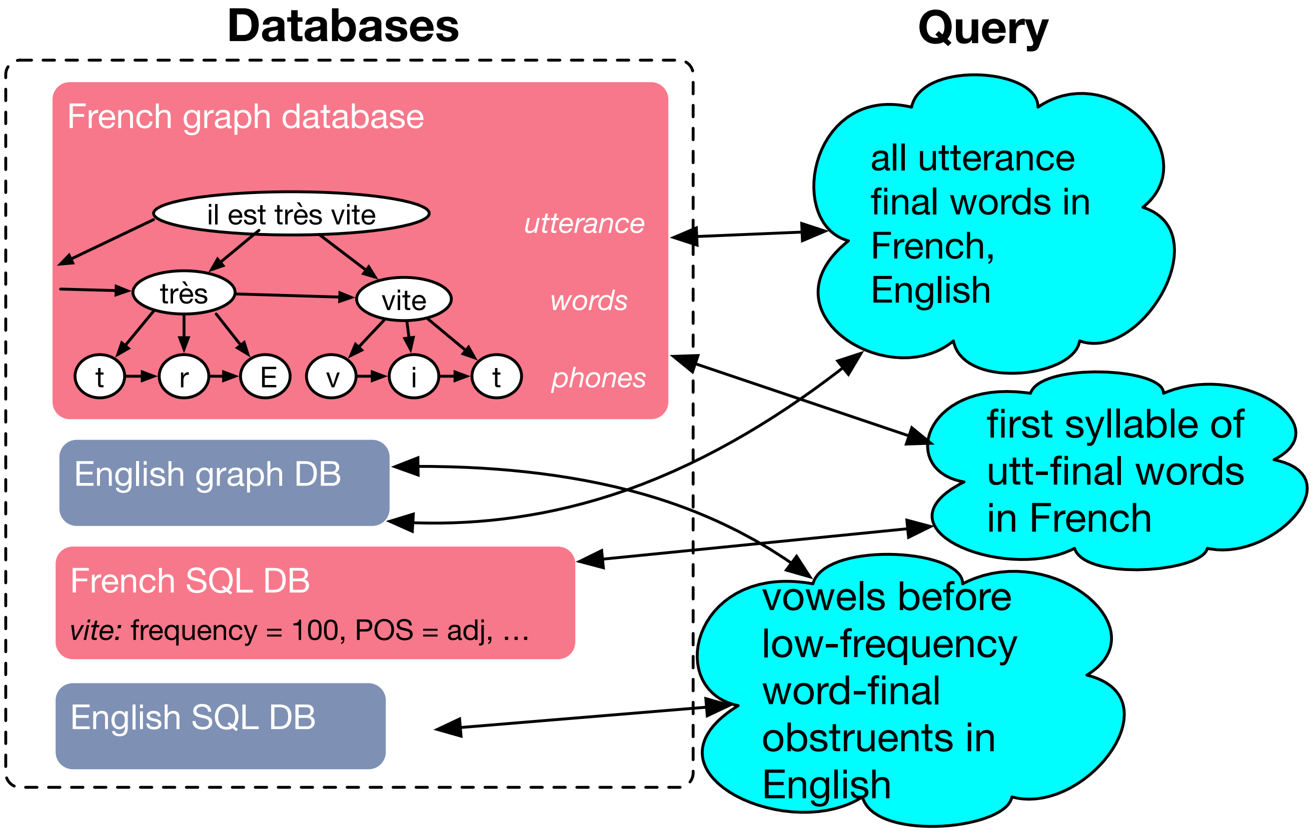
SCT: representation & enrichment

- DBs: contains properties of objects, relationships between them:
  - Positional:
    - Ex: Utterance position
  - Hierarchical
    - Ex: containing word
  - Temporal
    - Begin, end, duration



- Enrich with additional information:
  - Suprasegmental: pauses, utterances, speech rate
  - Acoustic: mean F0, formants, intensity

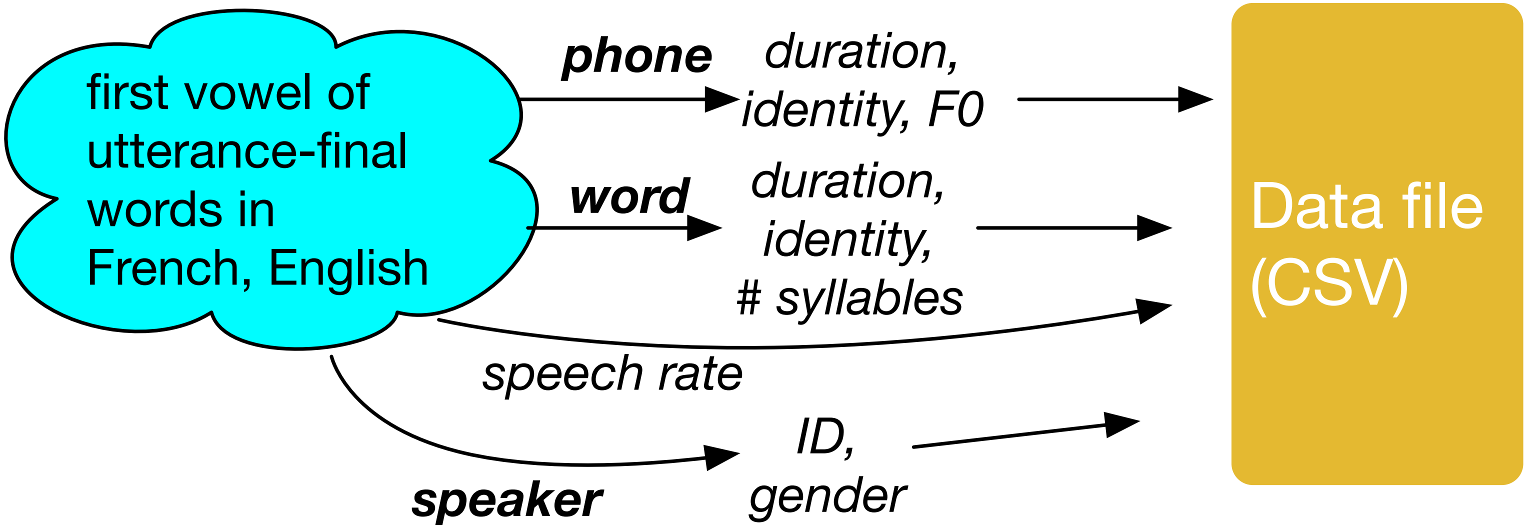
SCT: query



Databases
Query
French graph database
il est très vite
utterance
très
vite
words
t
r
E
v
i
t
phones
English graph DB
French SQL DB
vite: frequency = 100, POS = adj, …
English SQL DB
all utterance final words in French, English
first syllable of utt-final words in French
vowels before low-frequency word-final obstruents in English

- Find subset of linguistic objects

SCT: export


first vowel of utterance-final words in French, English

phone

duration, identity, F0

word

duration, identity, # syllables

speech rate

speaker

ID, gender

Data file (CSV)

- Properties of objects → spreadsheet
  - (→ R, Excel)

Case study

- Menzerath's Law (Menzerath, 1928, 1954)
  - Segments/syllables are shorter in longer words, in terms of:
  - duration per unit
  - # units (segments/syllable)

Overlapping

- Related: polysyllabic shortening
  - Syllable/V durations shorter in bigger words/prosodic domains
  - Ex: *stick*, *sticky*, *stickiness* (Lehiste, 1972)

- Cover term: duration compression effects

Duration compression effects

- Unclear: are DCE's
  - Universal?
  - Restricted to accented syllables?
  - Ex: Finnish, English, German
    (Siddins et al., 2014; Suomi, 2007; White & Turk, 2010)

- Our Q1: can we observe duration compression effects across typologically-diverse languages?

Duration compression effects

- Confounds:
  1. Accentual lengthening
  2. Domain-initial strengthening
  3. Word/phrase-final lengthening

(e.g. Sluijter, 1995; Fougeron & Keating, 1997; Oller, 1973; Klatt, 1973, 1975)

- Claim: maybe some of these things can be reduced to others
  - Ex: PSS is #1 or #3 (White & Turk, 2010; Windmann et al., 2015)
- Our Q2: can duration compression effects be reduced to a single other factor (across langs)?

Data

- Read sentences
- GlobalPhone (Schultz et al., 2013)
  - ~15 hours, 100 speakers / language
  - Czech, French, German, Polish, Russian, Swedish Hausa, Korean, Mandarin, Swahili, Turkish
  - Format: force-aligned TextGrids
- TIMIT (Garofolo et al., 1993)
  - 5.4 hours, 630 speakers, English
  - Format: text files

Import into SCT database:
TextGrid, TIMIT **importers**

Custom Kaldi aligner
(Povey et al., 2011)

One aligner/language; speaker-adapted triphone models

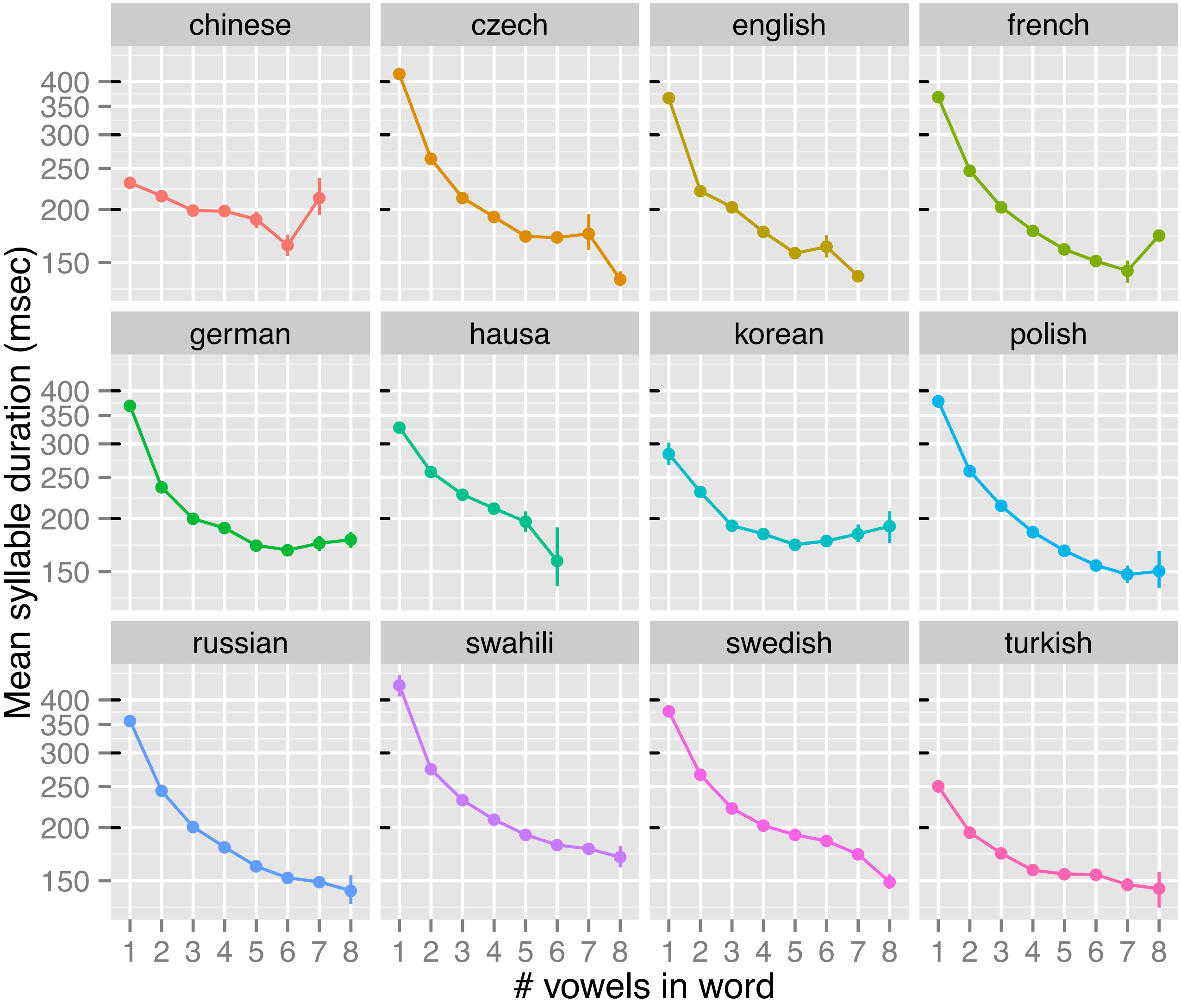
Procedure

- SCT query
  - Find: utterance-final words (>500 msec pause)
  - Export: # syllables, initial V duration, word duration (etc.)
- How does:
  - Mean syllable duration
  - Initial, final vowel duration
- Depend on:
  - Word length (# vowels)

    ?

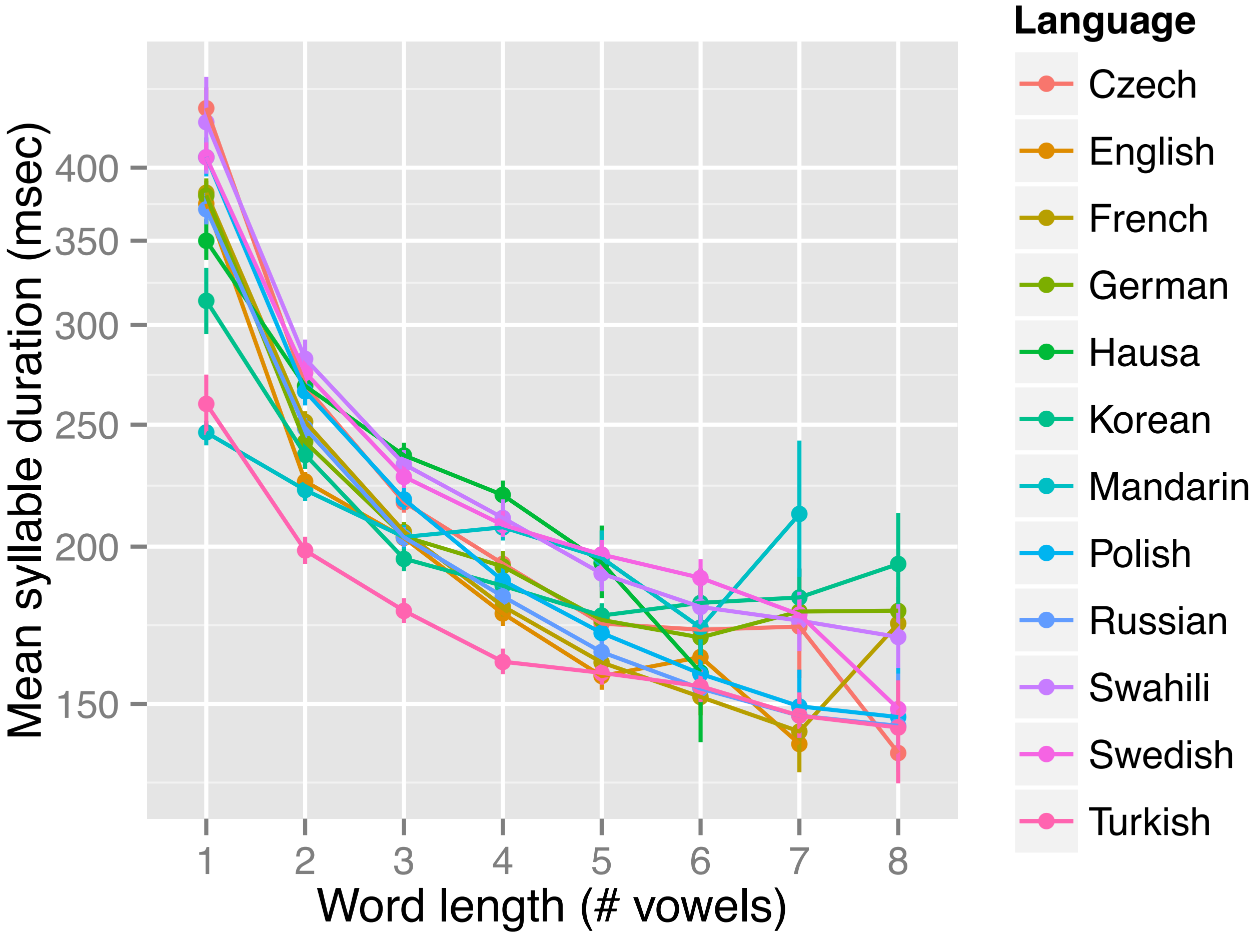
Proxy for syllables

Results: mean syllable duration



Compression effect across all languages

Results: mean syllable duration



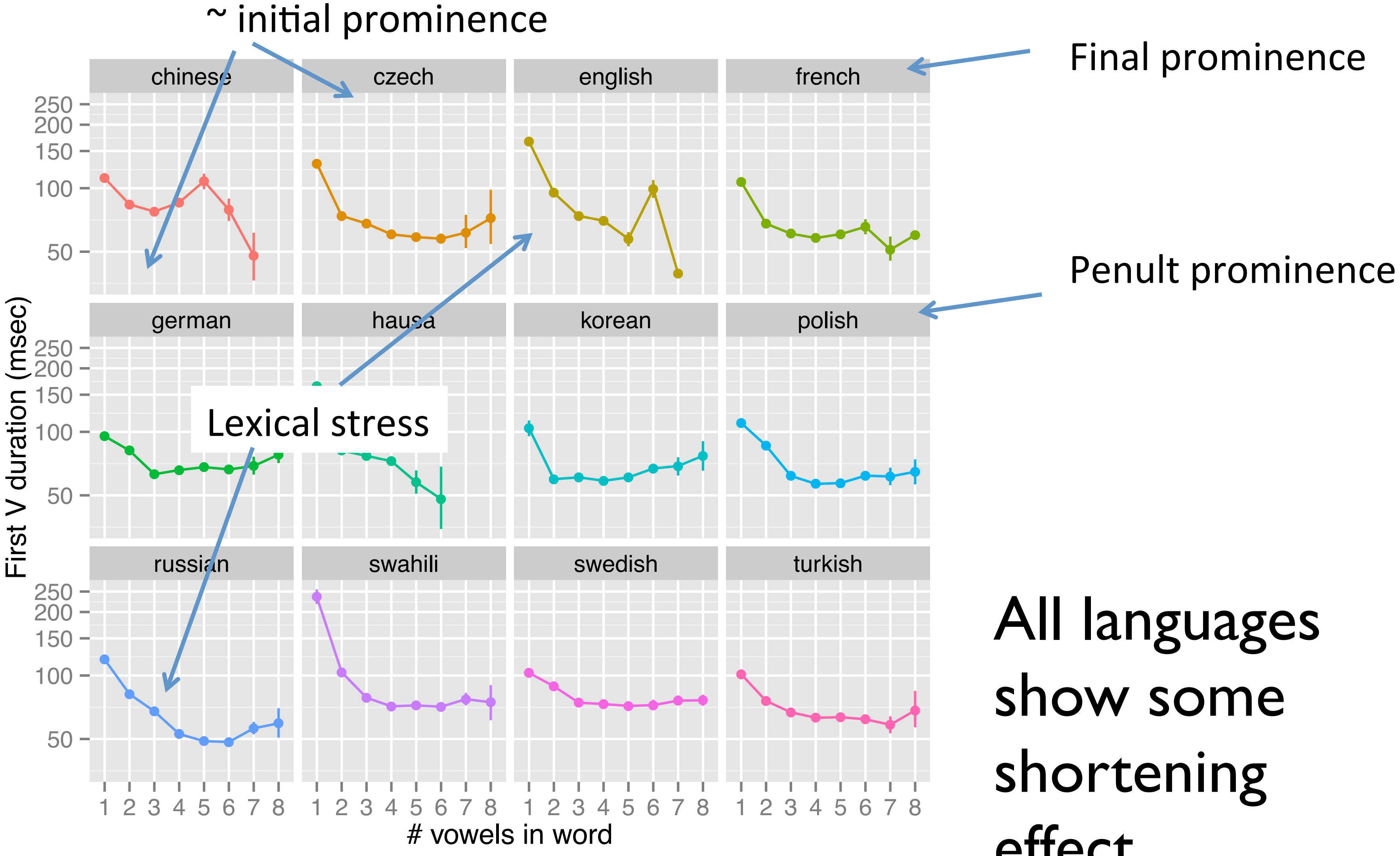
Very similar across languages!

Results: mean syllable duration

- Confounds: effect due to
  - Accentual lengthening (White & Turk, 2010)
  - PSS on stressed syll only?
  - Initial strengthening
  - Final lengthening (Windmann et al., 2015)

    ?

Results: initial vowel duration



All languages show some shortening effect

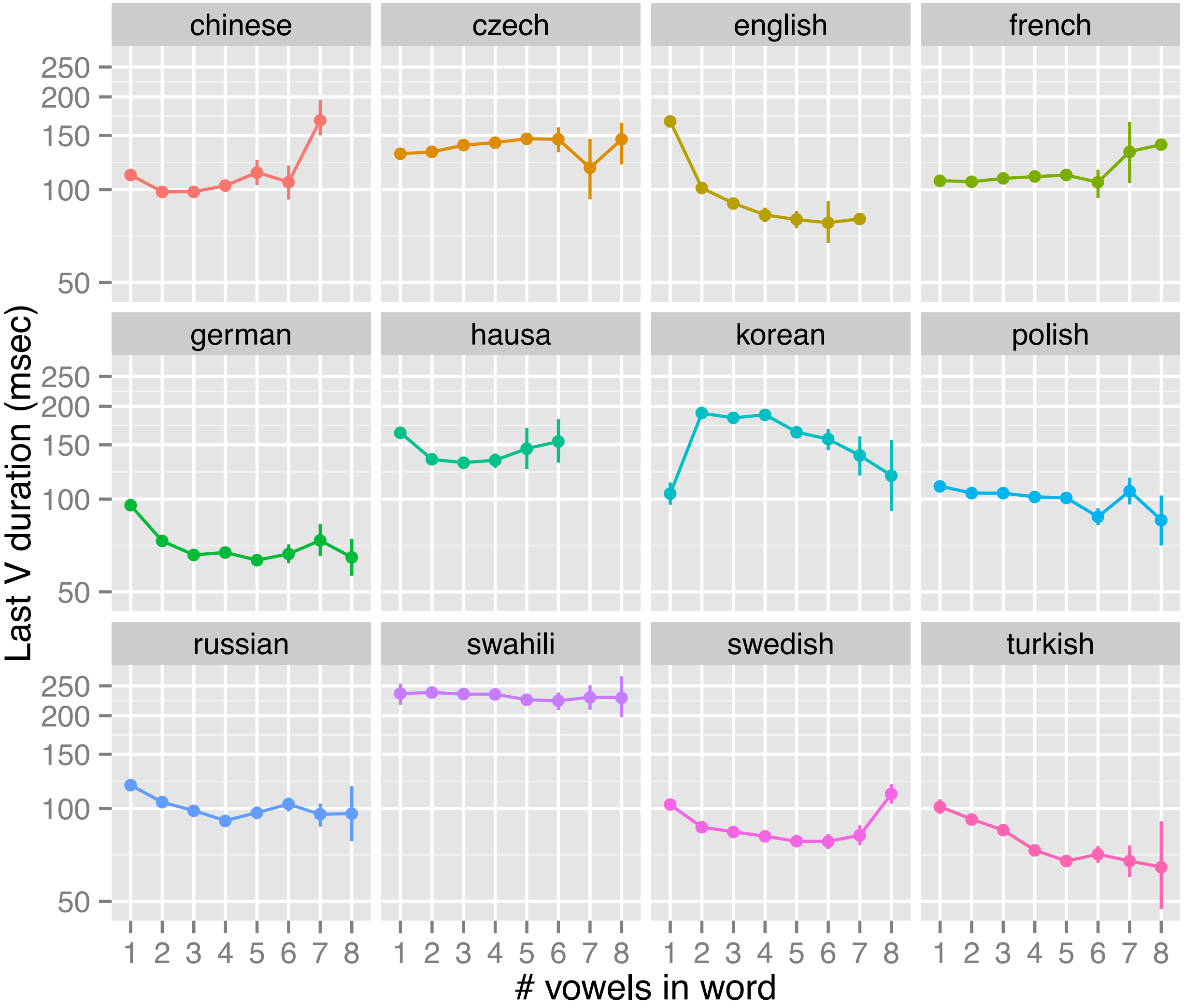
Results: initial syllable duration

- Consistent compression effect
  - (at least: 1-3 syllables)

- Very different prosodic systems

- Can’t be just
  - Accentual lengthening
  - Initial strengthening
  - PSS on accented syllables only

Results: final vowel duration



No consistent compression effect

Overridden by final lengthening + prosody?

(language-specific)

Summary

- Speech Corpus Tools:
  - Integrate large speech datasets, different formats
  - Query across them
- Goal: easy corpus studies
  - Find a set of objects
  - Export info about them
  - Make plots / do stat analysis
- Case study: duration compression effects may be
  - Universal
  - Not reduceable to (some) other effects

Thanks

- Montreal Language Modeling Lab members
- Funding:

INNOVATION.CA

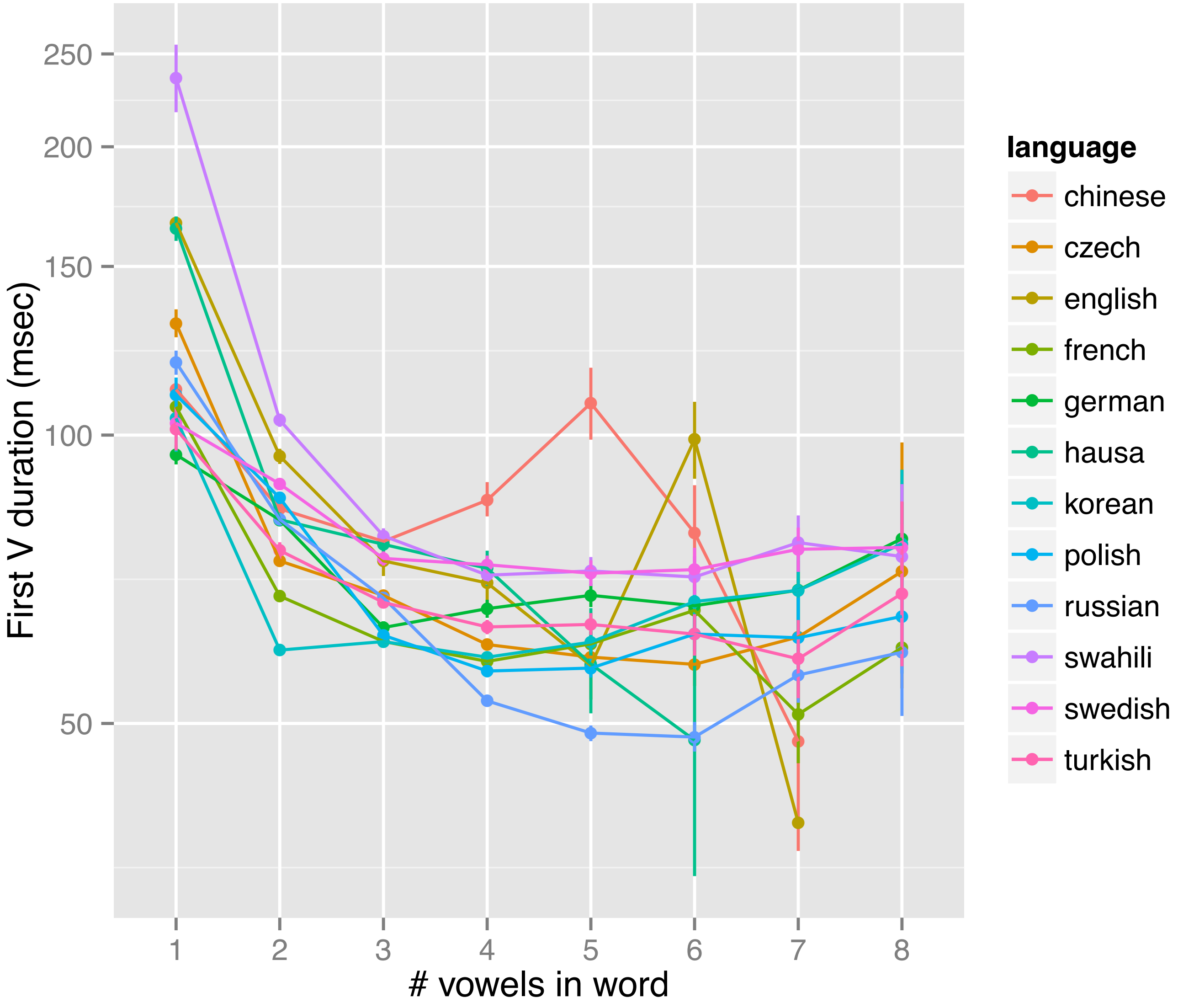
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

Fonds de recherche
Société et culture
Québec

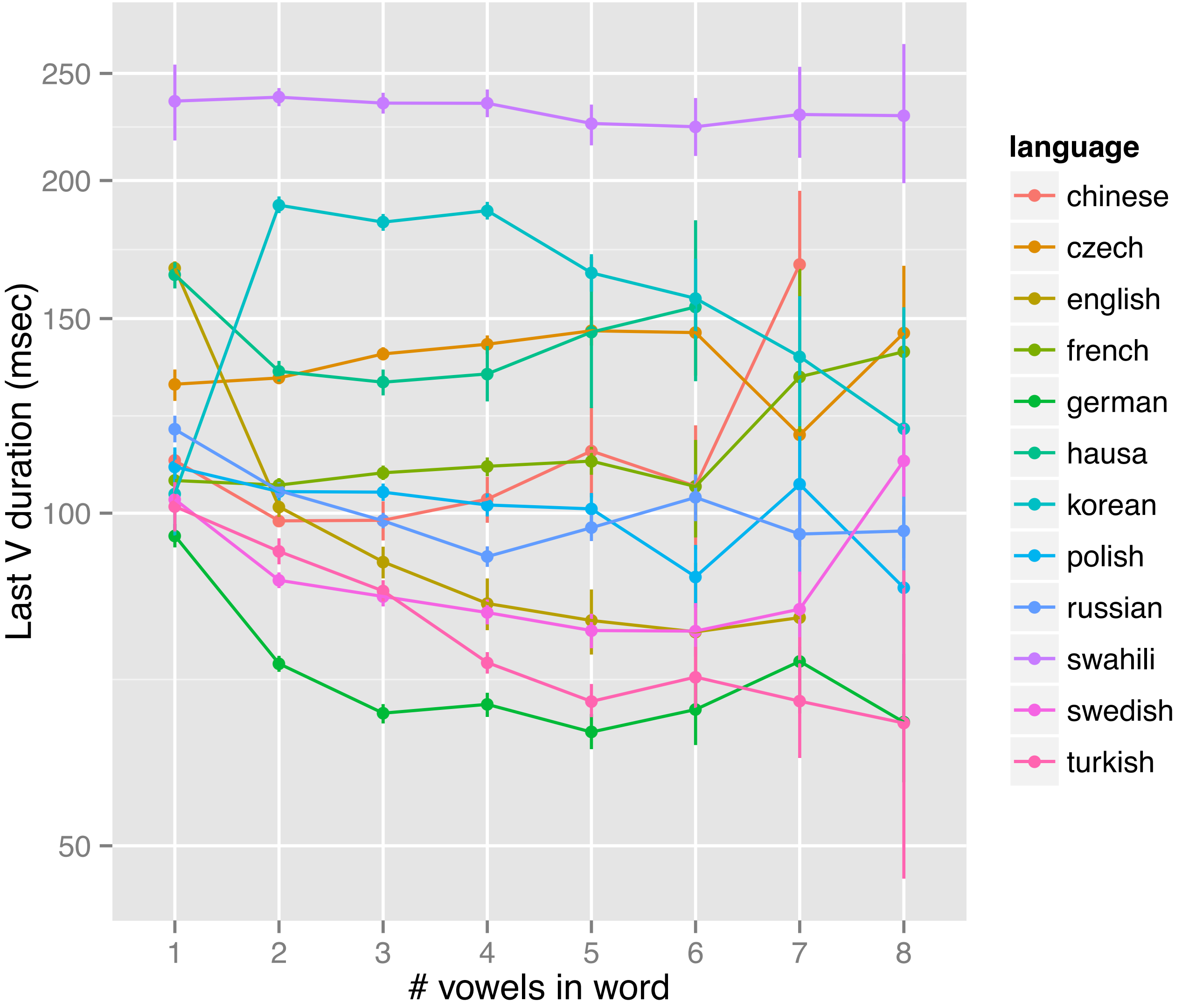
SSHRC CRSH

Questions

Results: first vowel duration



Results: final vowel duration


Last V duration (msec)
# vowels in word
language
chinese
czech
english
french
german
hausa
korean
polish
russian
swahili
swedish
turkish