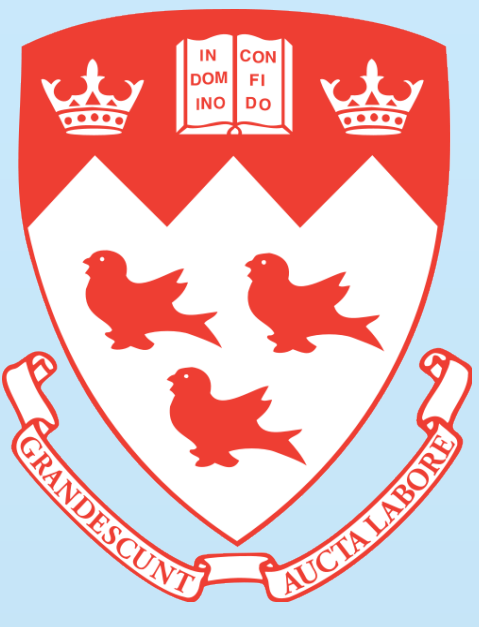


A system for unified corpus analysis, applied to duration compression effects across 12 languages



Michael McAuliffe, Morgan Sonderegger, Michael Wagner (McGill University)

1. Introduction

- Huge amount of annotated speech data exists
 - Corpora, labs, web
 - **great scientific potential**
- Requires software for **unified corpus analysis**
 - Integrating speech datasets
 - Querying across them
- Barriers:
 - Speech corpora are large, complex, diverse
 - technical, cost, access
- **Today:**
- Software: **Speech Corpus Tools**
 - Goals:
 - Minimal technical skill
 - Scalable, fast
 - Linguistically intelligent
 - ⇒ **easy(er), large-scale corpus studies**
- Case study / **proof of concept:**
 - Outstanding questions about duration compression patterns
 - **Large-scale unified corpus analysis can work, give new insights.**

2. Speech Corpus Tools

(Related: *Phon, EMU, LabCatt*: Fromont & Hay, 2012; Harrington, 2010; Rose et al., 2007)

Enrichment

- Add additional information to database
- Speech rate, F0, pauses, word frequencies, ...
- Standardized, customizable

Query

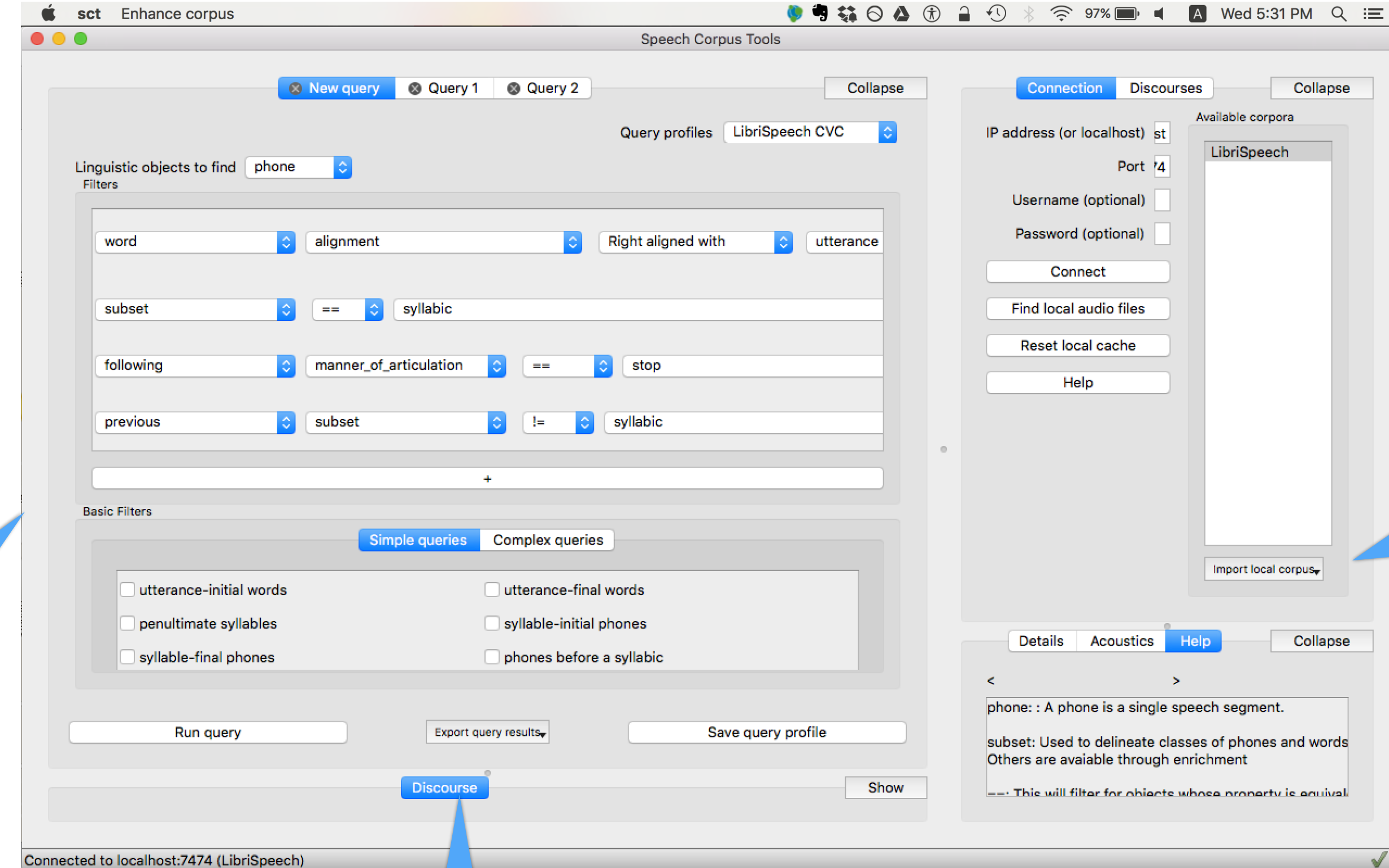
- Find linguistic objects
- Positional, hierarchical, temporal info

Export

- Properties of objects → spreadsheet

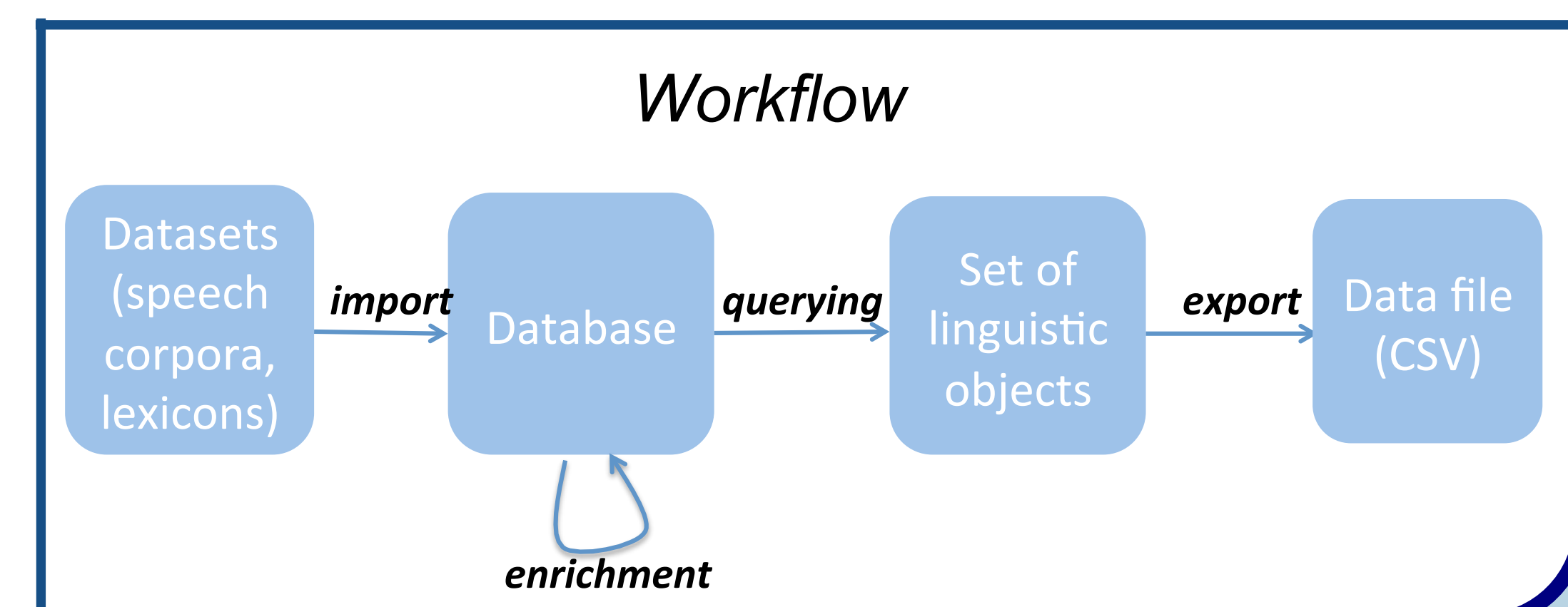
Import

- Datasets → standardized database format
- Abstract away from dataset format



Graphical interface + Python package

github.com/MontrealCorpusTools/speechcorpustools



3. Case study

- “Duration compression effects”
- 1. **Menzerath’s law** (Menzerath, 1928, 1954): Segments/syllables shorter in longer words, in terms of:
 - duration per unit
 - # units (segments/syllable)
 - Ex: *patisserie, pâté and pâte* (Grégoire, 1899)
- 2. **Polysyllabic shortening**: Syllable/V durations shorter in bigger words/prosodic domains
 - Ex: *stick, sticky, stickiness* (Lehiste, 1972)
- Still unclear about DCEs:
 - **Are they universal?**
 - Finnish doesn't show PSS (Suomi, 2007)
 - Is there only PSS but no Menzerath’s Law?
 - No DCE in unaccented syllables (Siddins et al, 2014; White & Turk, 2010)
 - **Q1: Can duration compression effects be observed across typologically diverse languages?**
 - **Reducible to confounds?** (e.g. Fougeron & Keating, 1997; Klatt, 1973, 1975; Oller, 1973; Sluijter, 1995)
 - Initial strengthening
 - Accentual lengthening
 - Final lengthening
 - **Q2: Can duration compression effects be reduced to a single other factor?**
- Q1, Q2: well-addressed by a large-scale study using SCT. Only need:
 - Durational info (syllables, words)
 - A lot of data per language

4. Data & Methods

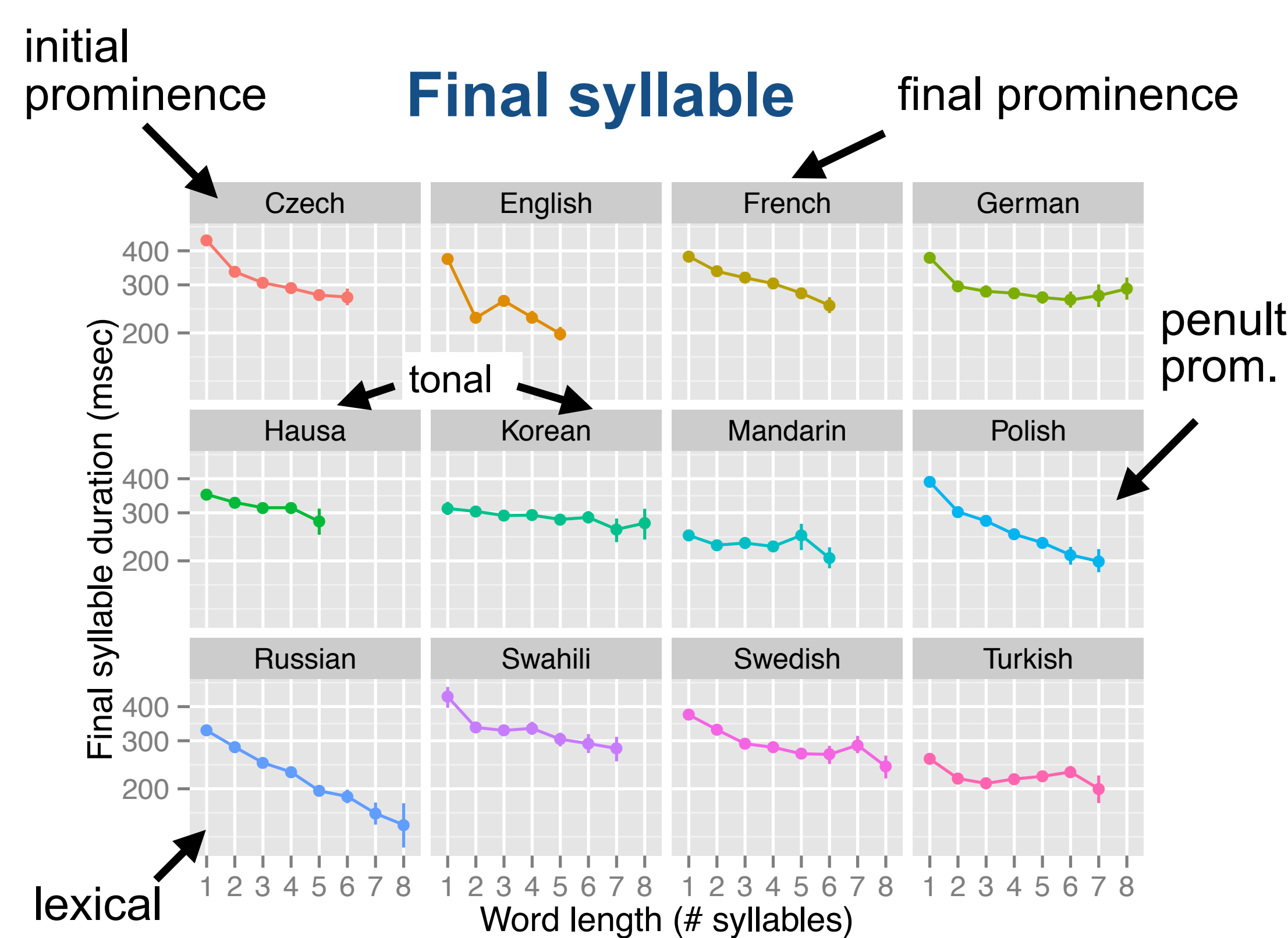
- Read sentences from 12 languages
 - English
 - Czech, French, German, Polish, Russian, Swedish, Hausa, Korean, Mandarin, Swahili, Turkish
- Process using SCT:
 - Import: TIMIT, TextGrids
 - Query: utterance-final words
 - Export: # sylls, word duration, duration of initial, medial, final syllable, etc.

TIMIT (text files, 5.4 hours, 630 speakers)

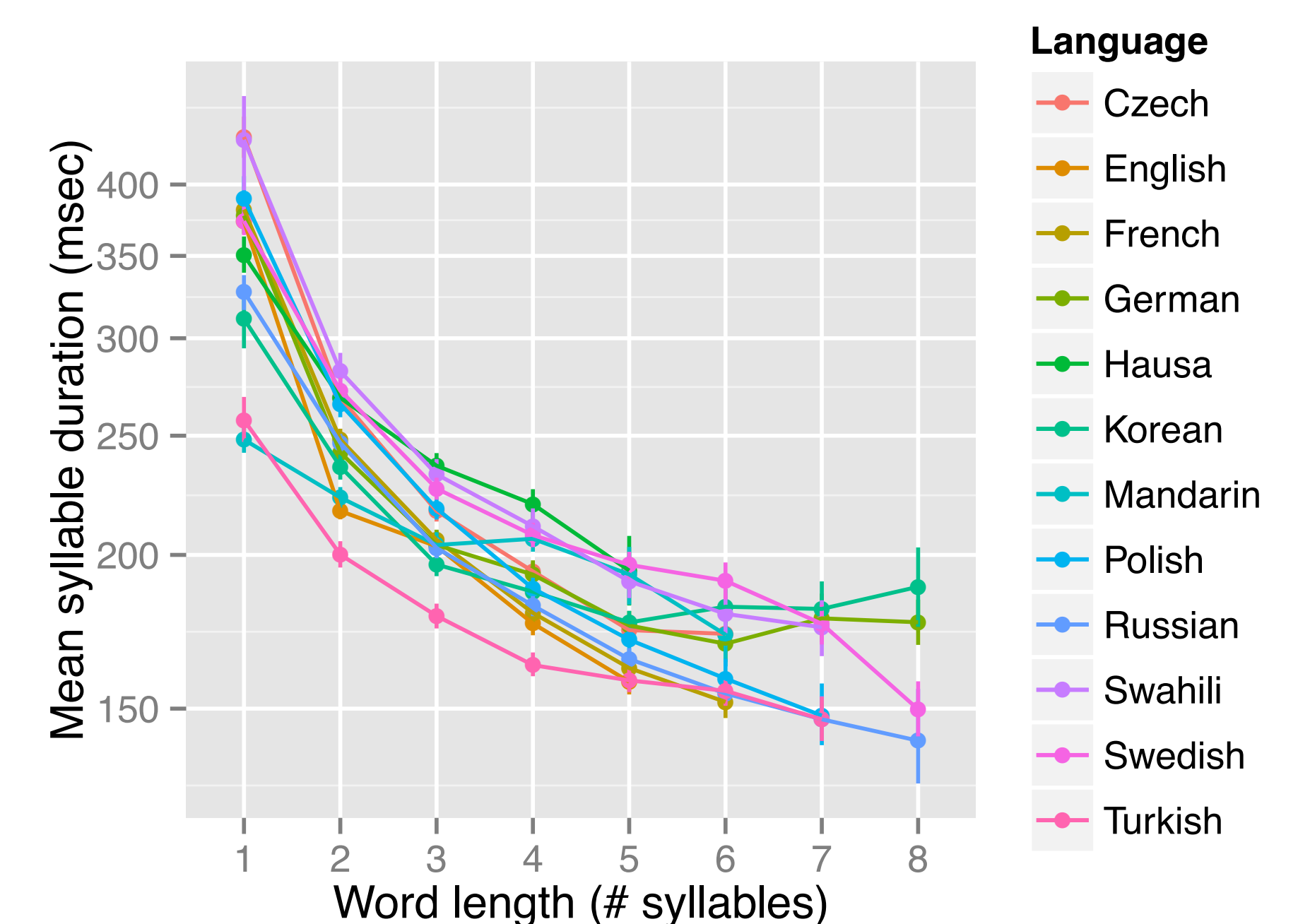
GlobalPhone (Force-aligned TextGrids, 15 h/~100 speakers per language)

Montreal Forced Aligner: trainable, Kaldi-based <https://github.com/MontrealCorpusTools/Montreal-Forced-Aligner>

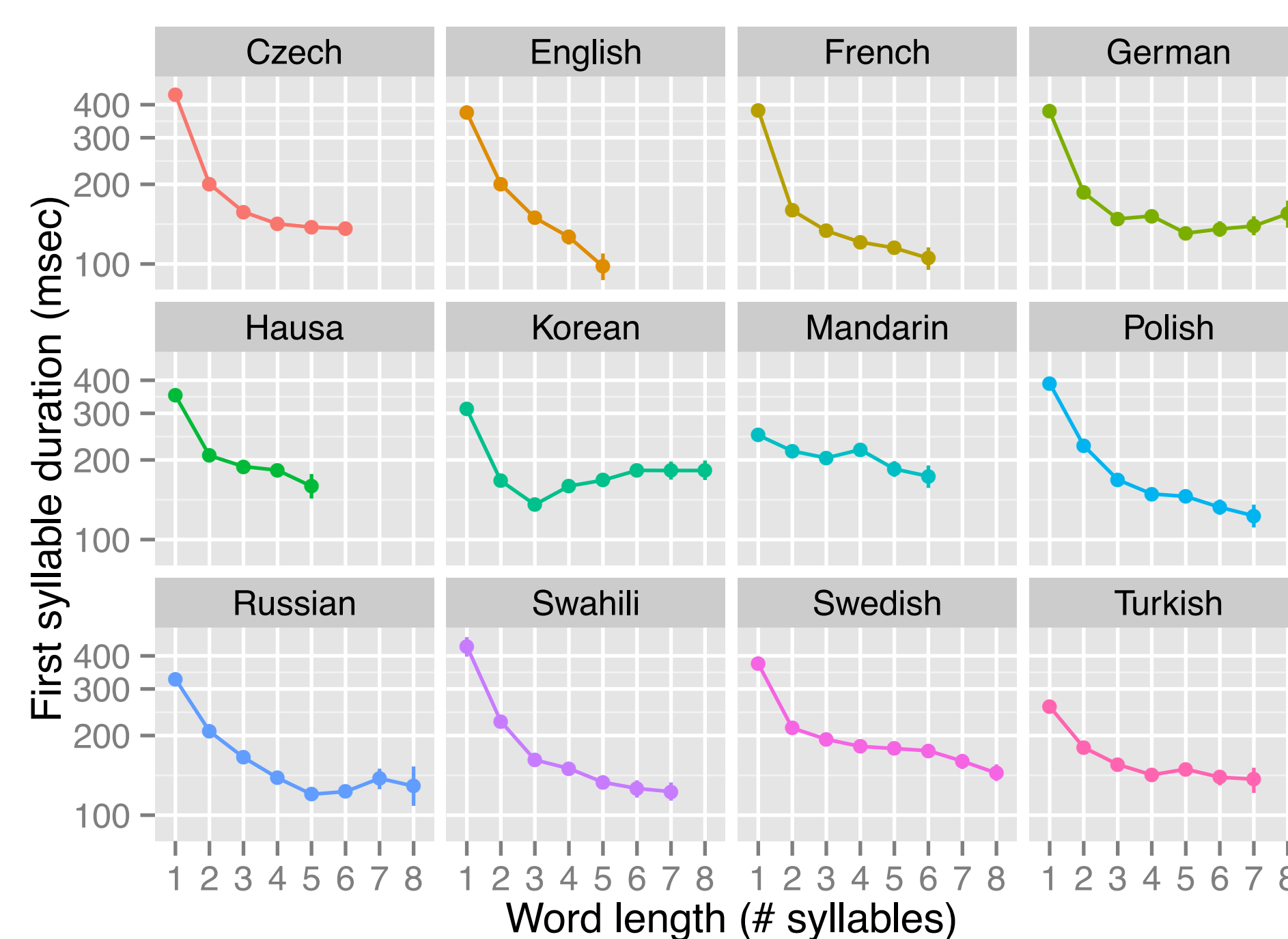
5. Results



Mean syllable duration



Initial syllable



- Mean: **Similar compression effect across all languages**
- Initial, final:
 - **Some compression effect across all languages**
 - very different prosodic systems!
 - differs by language (e.g. Ma, Ko vs. Fr, Po)
 - **compression not reducible to accentual lengthening** (cf. Fr & Po initial, Cz final)
 - And thus, Menzerath’s Law not reducible to PSS.
- Initial: **compression not reducible to final lengthening or initial strengthening**

Plots: mean and 95% CIs across mean durations for each speaker/word pair. Data restricted to #syllables with >= 25 word types.

Acknowledgements

- **Programming:** Elias Stengel-Eskin, Michaela Socolof
- **Funding:** SSHRC #430-2014-00018, #435-2014-1504; CFI #32451

References (selected)

Bird, S. & Liberman, M. (2001). A formal framework for linguistic annotation. *Speech Commun.* 33: 23-60. / Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *JASA* 51: 2018-2024. / Menzerath, P. and de Oleyza, J. M. (1928). Spanische Lautdauer: Eine experimentelle Untersuchung. / Povey, D. et al. (2016) <http://kaldi-asr.org/> / Schultz, T. et al. (2013). GlobalPhone: A multilingual text & speech database in 20 languages. In *Proc. ICASSP 2013* / Siddins, J. et al. (2013). The influence of accentuation and polysyllabicity on compensatory shortening in German. In *Proc. Interspeech*. / Suomi, K. (2007). On the tonal and temporal domains of accent in Finnish. *J. Phon.* 35(1): 40-55. / Garofalo, J. et al. (1993). *TIMIT*. Philadelphia: LDC / Windmann, A. et al. (2015). Polysyllabic shortening and word-final lengthening in English. *Proc. Interspeech*.