The Impacts of Modern Empire and the Languages of the American South

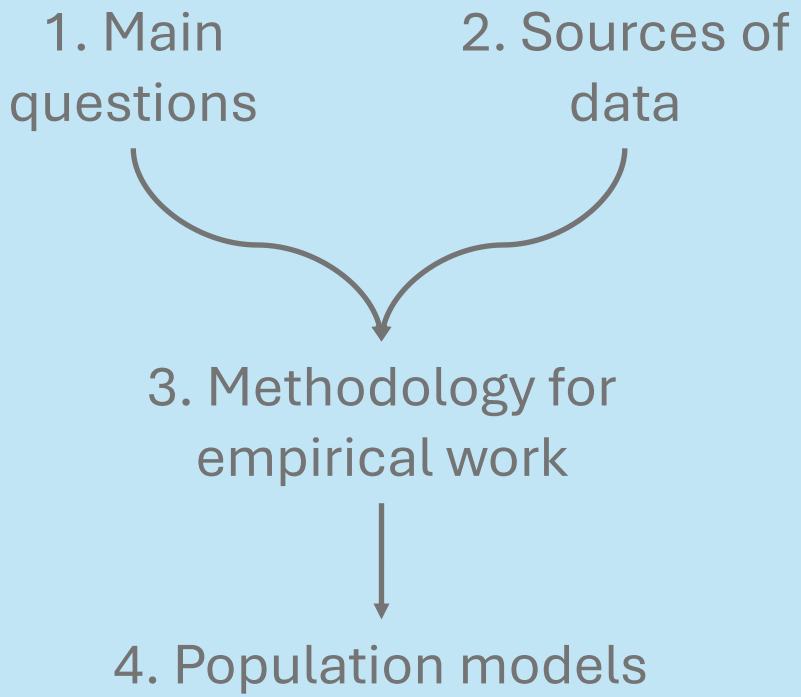
Seth (Dani) Katenkamp

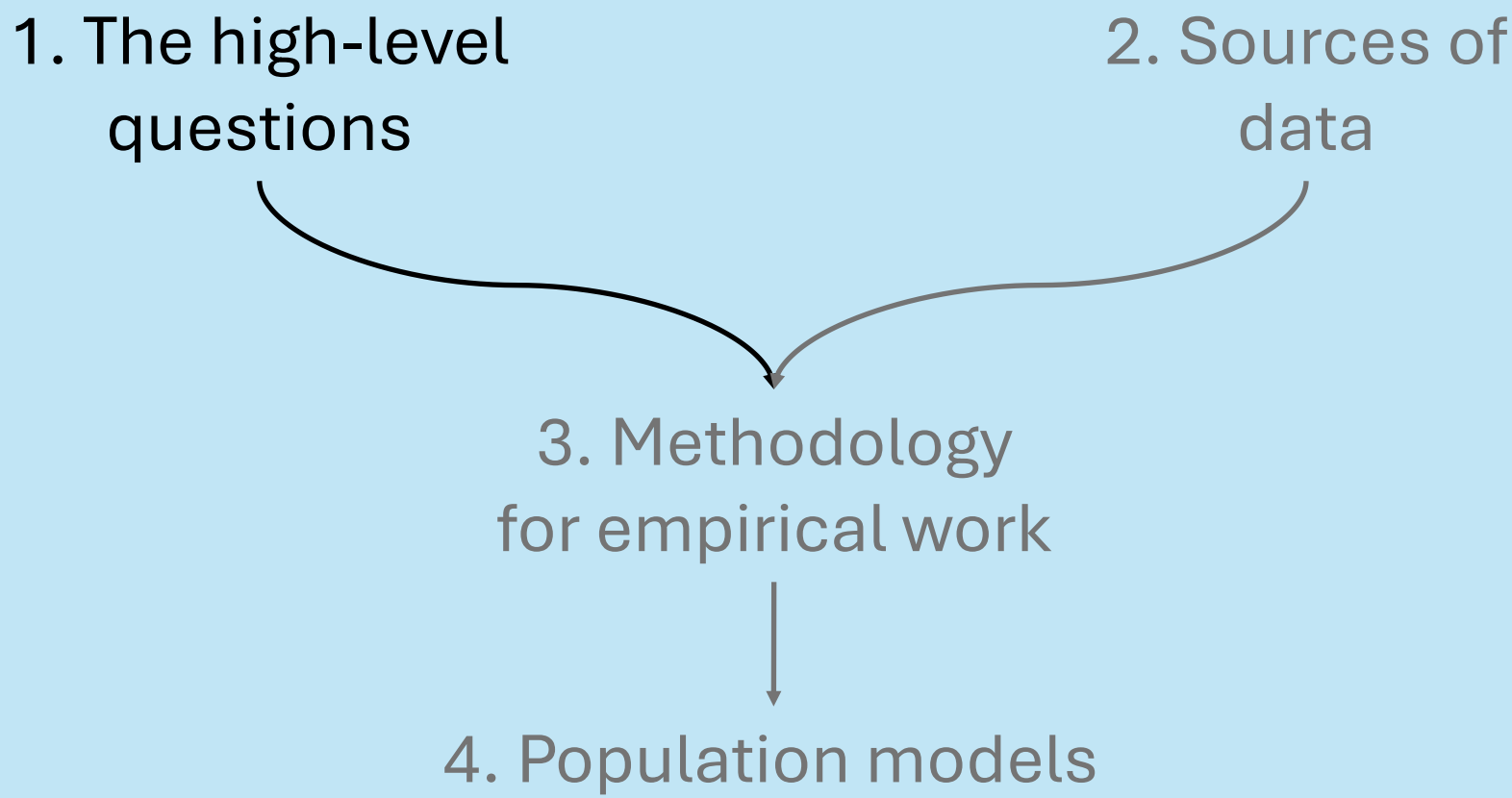
Committee: Claire Bowern, Jason Shaw, Alejandra Dubcovsky, and Jack Martin

(07 May 2025)

Roadmap

- 1. Introduction (main questions)
  - Contact in industrialized empires, variable outcomes
- 2. Sources of Data
  - Prepping corpora
  - Choctaw
- 3. Methodology for empirical work
  - Variables
  - Interpretation
- 4. Population models
- 5. The plan




1. The high-level questions
2. Sources of data
3. Methodology for empirical work
4. Population models

Language contact

- When different languages come into contact, a variety of language change phenomena occur:
  - Borrowing, neologization, morphosyntactic restructuring, language shift, creolization
- Does a unique contact scenario → unique language change?
- Specifically: **What types of change do we expect in a contact ecology characterized by the dynamics of modern colonialism?**

Dynamics of modern colonialism

- Global empire → attempts at genocide and erasure
- Capitalism → conflicting ideologies of trade/value (Galloway, 2009)
- Industrialization → devaluing of traditional lifeways, disrupting restorative practices re ecosystem
  - Replacing previous methods of sustenance with collaborative mass production
  - New trades, fields of expertise
- Urbanization → displacement from traditional environment, mixing of communities (Oakland, Los Angeles, Oklahoma City, etc.)
- Mass media → increased/ubiquitous exposure to dominant language
  - Initially paired with a lack of access to that communication infrastructure for their own language

Expectations/Intuitions

- General inundation, combined with socioeconomic incentives lead to greater proficiency in colonial L2, and more code switching
  - More borrowing
  - Resort to hypernyms to fill gaps created by the loss of infrequent lexical items
  - General language shift
- Altered set of appropriate domains for language use
  - Non-native spatial domains: corporate environments, industrial workplaces, urban areas, locations of forced migration, etc.
  - Cessation of traditional lifeways, especially ceremonies and material culture

Complicating those expectations

- Counter-examples:
  - Shift to polysynthesis in Bardi in the 20th century (Bowern, 2012)
  - Borrowing prevented by language-mixing taboos (Epps, 2009)
    - Specific attitudinal factors prevent the expected change!
- Localized effects:
  - Color with particular referents
  - Borrowing prohibited in particular social domains, e.g. ceremony
    - Prevents attrition or replacement of native vocabulary in that part of the lexicon
  - Less frequent use of borrowings in social elites (Katenkamp, 2025)
- Revitalization/reclamation/resistance movements
  - Young people's varieties, mixed languages, creoles complicate narratives of attrition
- So contact-based change
  - Isn't always either attrition or shift (better characterized as a variety of changes)
  - What happens is often determined partly by societal attitudes

The big questions

a) How do speakers of Indigenous languages change their behavior in response to modern (industrial and imperial) history?

b) What sorts of domain specificity (speaker subgroup, speech domain, region of the lexicon, etc.) do we see in these changes?

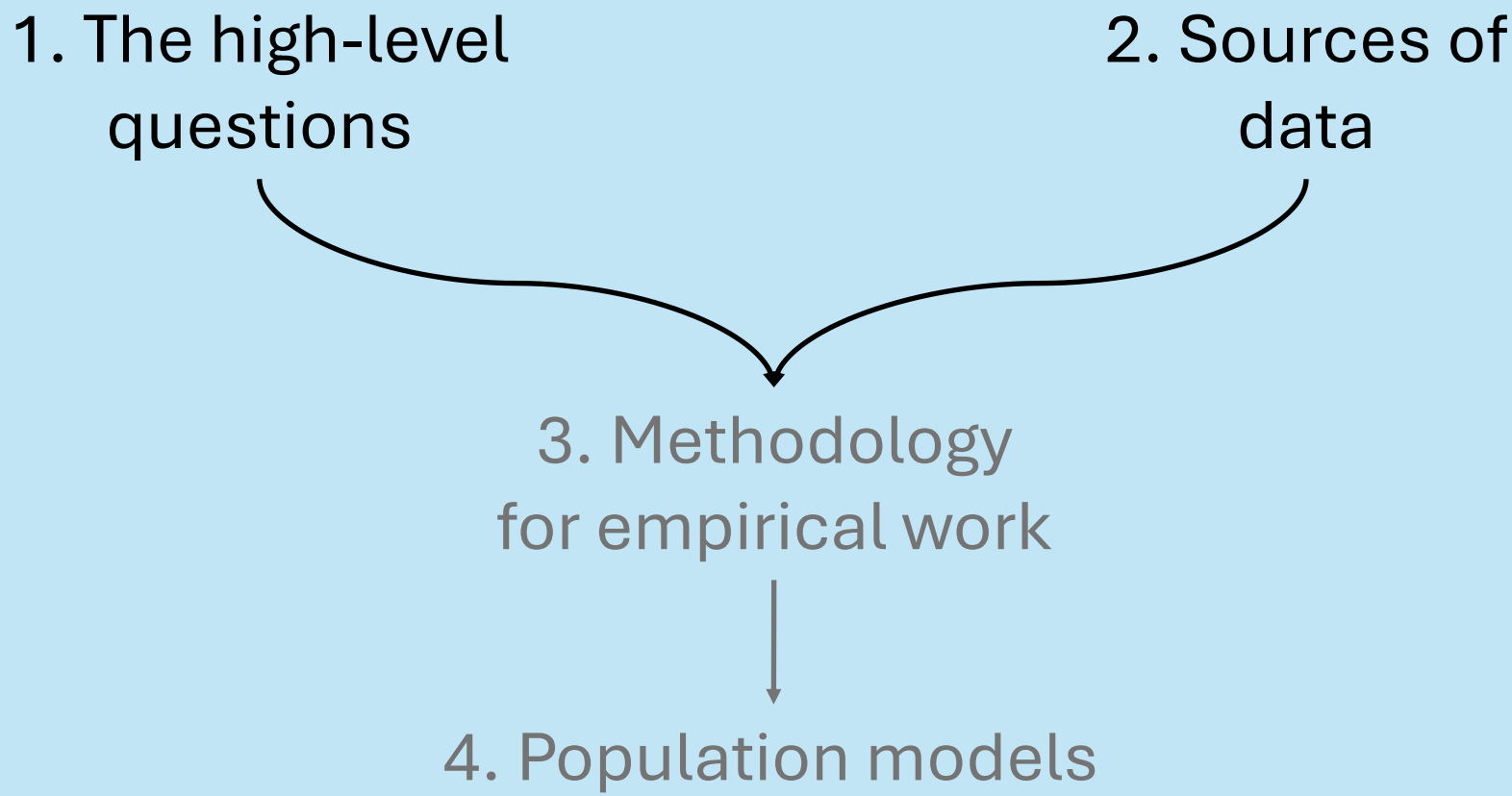
c) How do these changes relate to individual sociopolitical events and community values, e.g. when do speakers assimilate versus dissimilate?

Narrowing the focus

- The contexts and trends relevant to those questions are too varied and complicated to explore at a global scale
  - cf. Bromham et al.'s (2022) study of likelihood of endangerment as a singular variable
- So we look at a few languages in the American South
- Extremes of colonialism:
  - Very old European-Native American contact
  - Shatter Zone period (1540-1715) involved extensive migration, coalescence of disparate groups, and population decline
  - Small pluralistic confederacies victims of American expansion and Removal to Oklahoma
- Indigenous groups here have similar social architecture and experiences of European contact, but there still exists cultural and historical variation between them.

Specific opportunities for observation

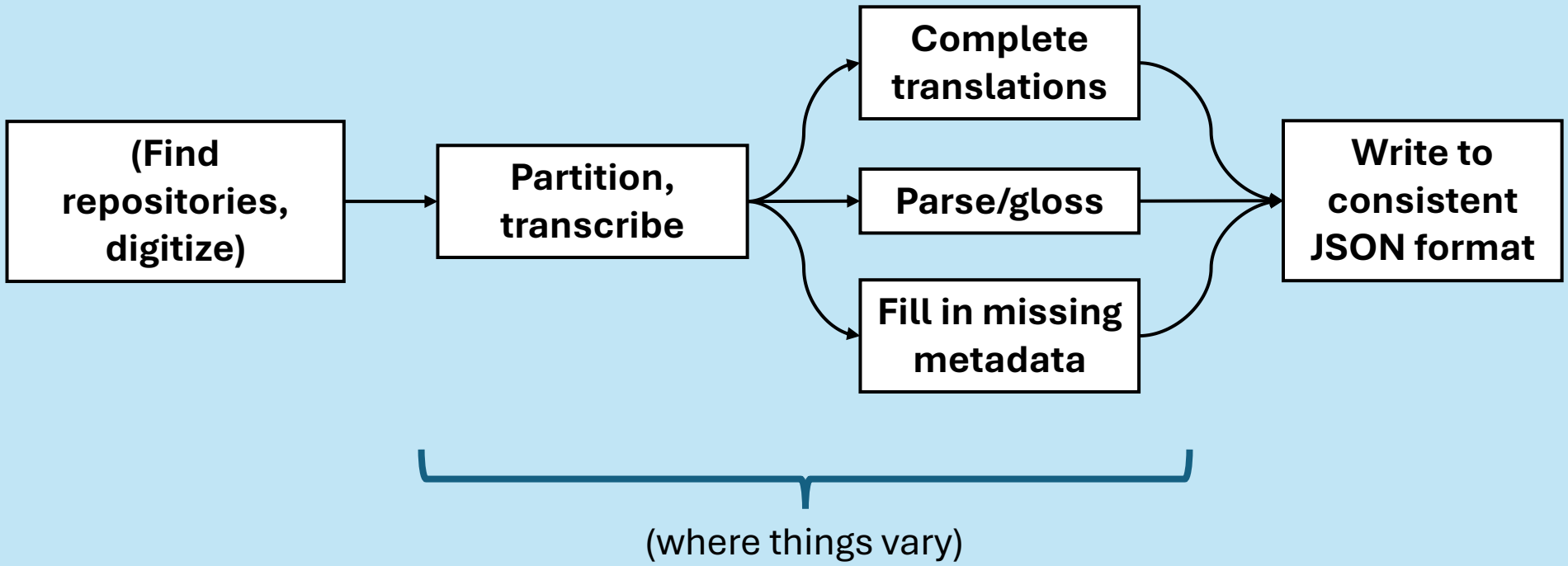
- Muskogean languages
- Earliest material is from the 18th century
  - 1740 for Mvskoke/Creek
  - 1775 for Choctaw
- Significant corpora from the 19th and early 20th centuries
- Modern documentation to compare to


1. The high-level questions
2. Sources of data
3. Methodology for empirical work
4. Population models

What exists?

- Lots of scattered written material, typically produced by Native people for Native people, from about 1740-1910
- Written material from later in the 20th century via language documentation (e.g. Haas, 1944, 1945)
- Some recordings from the 21st century (also language documentation)
- Mainly Choctaw, Creek, Koasati
  - 3/4 branches of the family
  - On the order of a few 100k words / each language
- (Ask me in the Q&A about specific materials/languages)

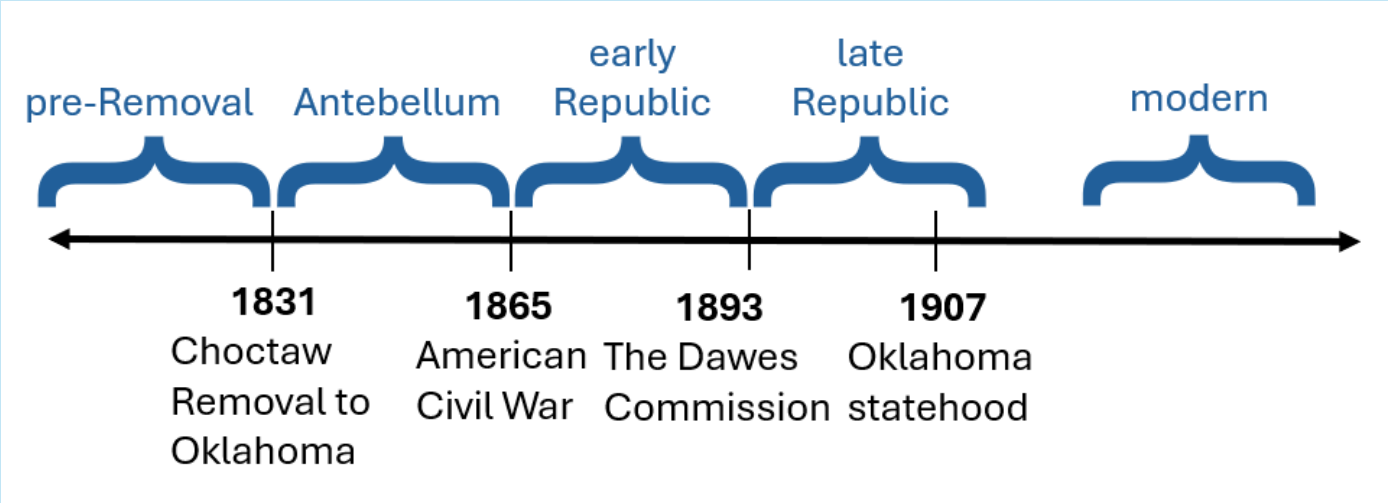
Basic (maximal) pipeline for prepping corpora



Choctaw

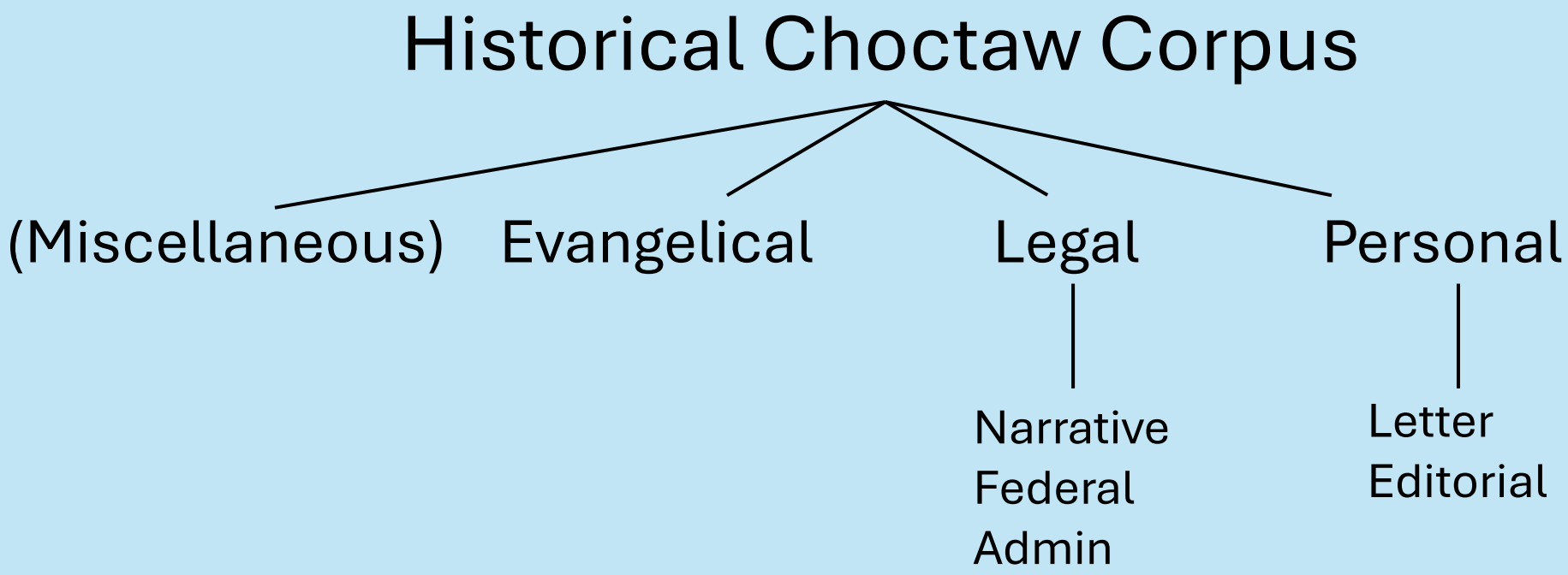
- Going to focus on Choctaw material (specifically 19th century), because that material is fully processed
- Basic facts about the Historical Choctaw Corpus:
  - 630k words
  - Several hundred speakers
  - Demographic variation: age, social class, race, gender

Sub-corpora



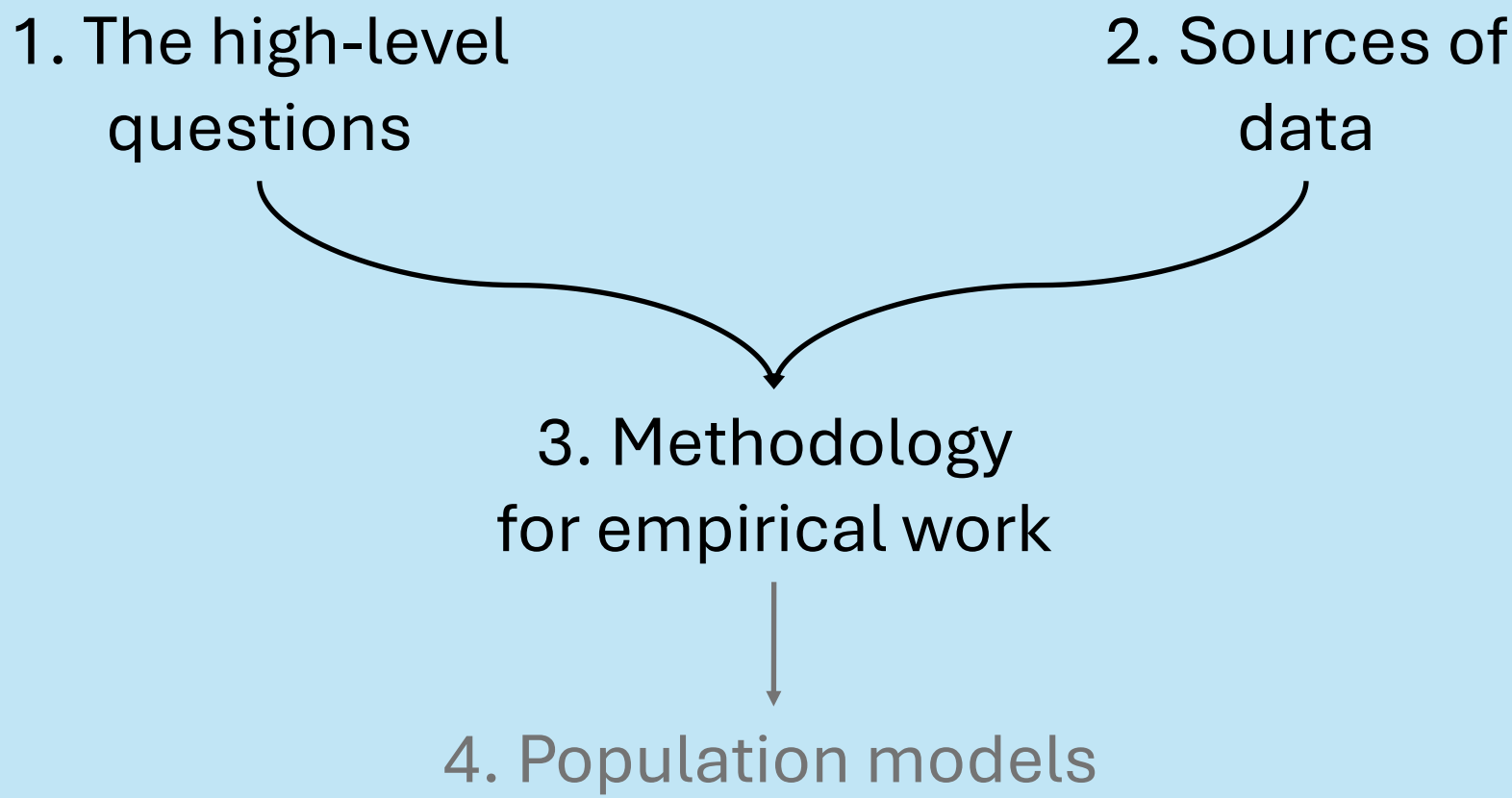
- Time periods chosen based on important socio-political events
  - The middle three intervals are similar in length (34, 28, 24 years)
- Other subdivisions that might be significant?
  - Genre (alt. macro-genre, which is more coarse-grained)
  - Speaker class
  - Medium (transcribed speech versus written composition)

Genres for Choctaw



Unevenly attested (number of utterances)

	evangelical	legal	personal	total
pre-Removal	1664	724	**000**	2388
Antebellum	3284	51	1753	5088
early Republic	43	14085	106	14234
late Republic	**000**	10122	214	10336
total	4991	24982	2073	32046


1. The high-level questions
2. Sources of data
3. Methodology for empirical work
4. Population models

Variables

- Meakins et al. (2019)
  - “language features” for which there are variant expressions
  - Variants may originate in one language or another, or be innovative
  - Lexical and morphosyntactic variables
  - “The 120 language features were **chosen because they vary rather than for their specific patterns of change** (such as simplification) in order not to bias the analysis.”
- Useful methodology because
  - We identify maximally local changes
  - Not as vulnerable to empirical gaps (cf. alternative methodologies for finding out if the total lexicon has shrunk)

Different types of variables

- Lexical vs. morphosyntactic

Borrowings:	‘donkey’	*issoba in haksobish falaaya* ‘like a deer, with long ears’ *mul*
Hypernymy:	‘wine’	*oka homi* ‘bitter water, liquor’ *panki okchi* ‘wine, grape juice’

Different types of variables (cont.)

Options for constructions: where one lives

(a) *(place) aayahanta*

aa- atta -hVn

loc- be.at -rep

‘be there (continually)’

(b) *(place) aayokchanya*

aa- okchaya -n

loc- live -dur

‘living there, being alive there’

General use of complex morphology: How consistently are coordinate clauses marked with switch reference?

Axes of variation

- Meakins et al. (2019) look at variants in a mixed language which can be tagged for two different properties:
  - (a) Gurindji, Kriol, or innovative origin
  - (b) simple or complex
- This allows the trends in the usage of variants for each variable to be identified as indicating shift along the axes of
  - (a) preference for forms from one of the source languages
  - (b) preference for simplicity
- Two possible axes for the Muskogean data: hypernymization and Anglicization

Hypernymization

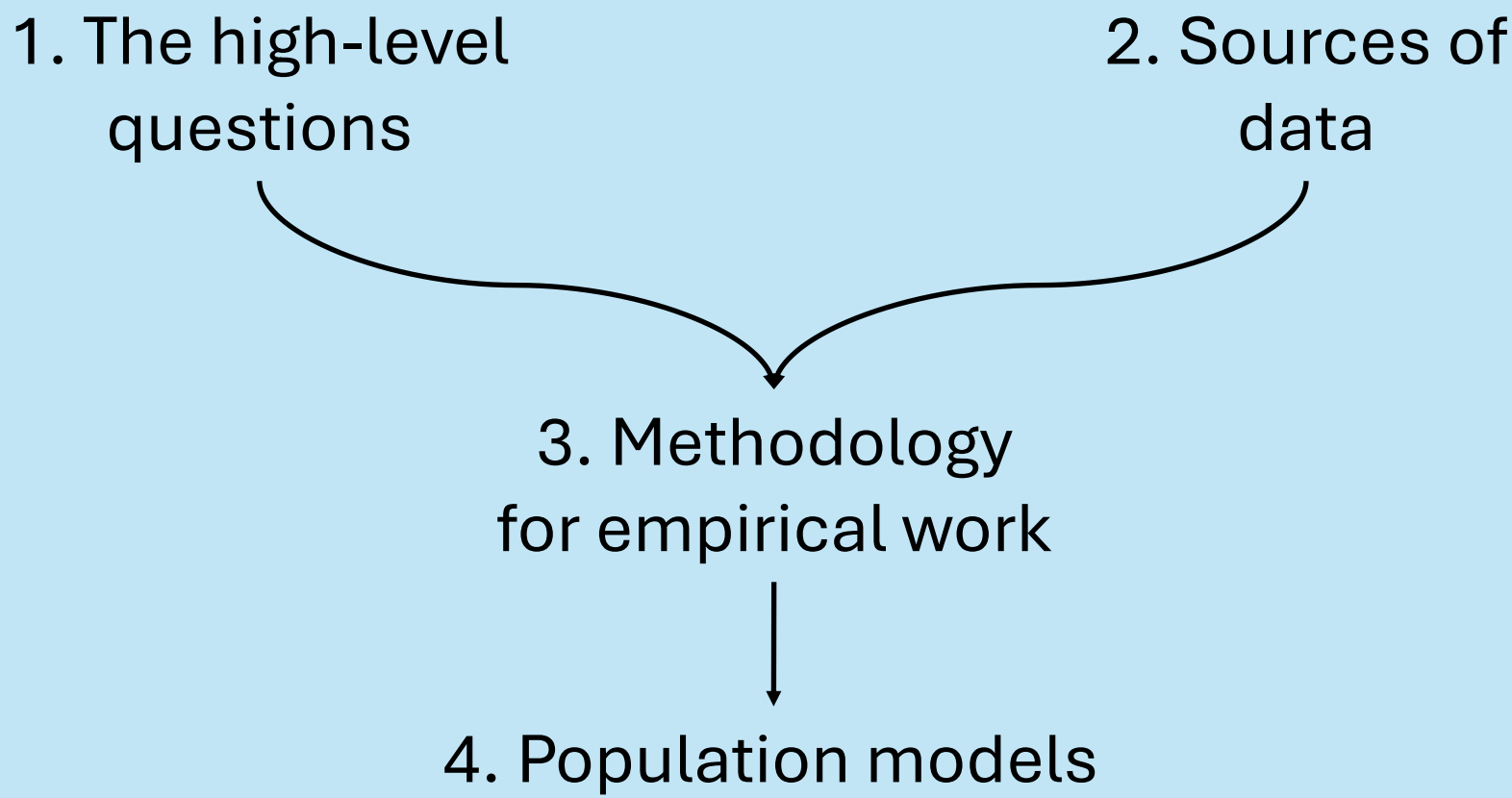
- “Simplicity” as a concept is ethically and formally fraught
  - Absolute complexity: “the number of elements…the amount of information needed to describe them,” e.g. conjugation classes, number of contrasts, etc. (Nichols, 2009:111)
  - An example from Meakins et al. (2019:294): “borrowing prepositions over case morphology”
- Hypernymy eschews metrics of complexity in the language system for more descriptive properties
  - Are there fewer lexical items in use in the language?
  - Intuitive transmission advantage to a smaller lexicon- but that doesn’t automatically mean that’s what happens

Anglicization

- Obviously full borrowings from English, but also semantic borrowings (Brown, 1999), e.g. *correr* for ‘run for office’ in Brazilian Portuguese
- Changes in the number of contrasts
  - Increase to match English: color terms
  - Decrease to match English: sibling terms, pluractional marking in verbs

Interpreting distribution

- Thinking about individual texts
  - So we don’t over-represent large texts
  - So we can think about the distribution of behavior across individuals in the population
- Two types of changes that can happen to the distribution of a given variable:
  - Change in average use of a variant (across the population)
  - Change in how widely distributed the individual behaviors are (stdev)
- Working out the social dynamics that produce the two types of change using a population model


1. The high-level questions
2. Sources of data
3. Methodology for empirical work
4. Population models

Modeling goals:

- Identifying specific parameters that generate the two types of change
  - A mixture of bias (for or against a variant), strength of conformity to different parts of the population, etc.
- By working out the parameters that generate a particular trajectory, we establish quantitatively the strength of the bias (and possibly the types of bias) active for that variable
  - Comparing to a null model, where the distribution is the result of non-selective drift
  - Establishes a likelihood of social selection

Overall structure of the model (for one variable)

- Population of constant size
- Individuals have a single value representing how frequently they use the target variant of the variable
  - Randomly assigned based on a truncated Gaussian distribution ([0:1]) centered on the mean of variant use across the population
  - So no separate phases of innovation and then spread
- Discrete timesteps: what happens?
  - Individuals are randomly paired together to ‘interact’ until all individuals have interacted at least once
  - Interaction changes the frequency of both individuals (see next slide)
  - After all interactions take place, the mean across the population is updated
  - Then some of the population is replaced (a “mortality rate” parameter)

‘Interaction’ between individuals

f **= speaker’s frequency of use (0 < f < 1)**

r **= speaker’s resistance to change**

b **= population general bias towards the variant**

μ **= mean frequency of use across population (0 < μ < 1)**

g **= interlocutor frequency of use (0 < g < 1)**

t **= timestep**

amount that the frequency will change

$$f_{t+1} = f_t + \frac{1}{r} \cdot \frac{1}{1 + e^{-b(0.5 - f_t)}} \cdot \left(-1(f_t - \mu) - 0.5(f_t - g)\right)$$

current frequency

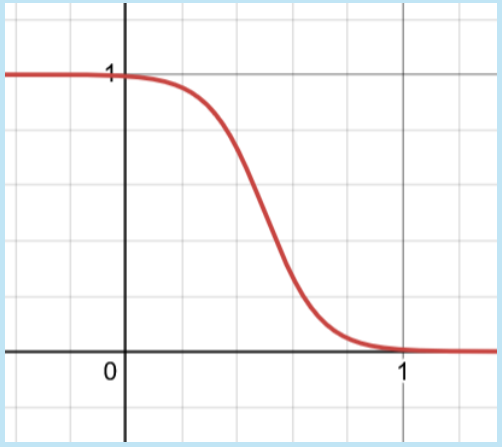
resistance

bias towards variant (sigmoidal)

contextual pull, a function of the speaker’s distance from the general mean and their interlocutor’s use

Some assumptions

- Speakers have some awareness of the general trend in their population, beyond the scope of their current interaction
  - Unlike models like Kirby (1998)
  - Speakers in the real world encounter their language ambiently
- Bias calculated via sigmoid function: the closer speakers are to whichever pole they're biased towards (0 or 1), the weaker the effect



Some assumptions to change in future versions of the model

- Currently the resistance to change is the same for all speakers
- No social subclasses
  - Segregation
  - Behavioral differences (e.g. propensity towards borrowing, Katenkamp, 2025)
- No selection for replacement based on age
- (these are all parameters which should be added)

Interpreting results

- Changes in mean = changes in the population's use of variant
- The *b* (bias) necessary to generate the shape of the curve (the change in m) show the degree to which the change in frequency could be the result of drift
  - Greater bias, less likely to be drift
- Distribution of the population: wide or narrow
  - Width in a given timestep isn't necessarily insightful
  - But *change* in width represents homogenization/heterogenization of the population's behavior (suggesting social salience)
- Changes in the model parameters mid-simulation in order to capture changes in trajectory

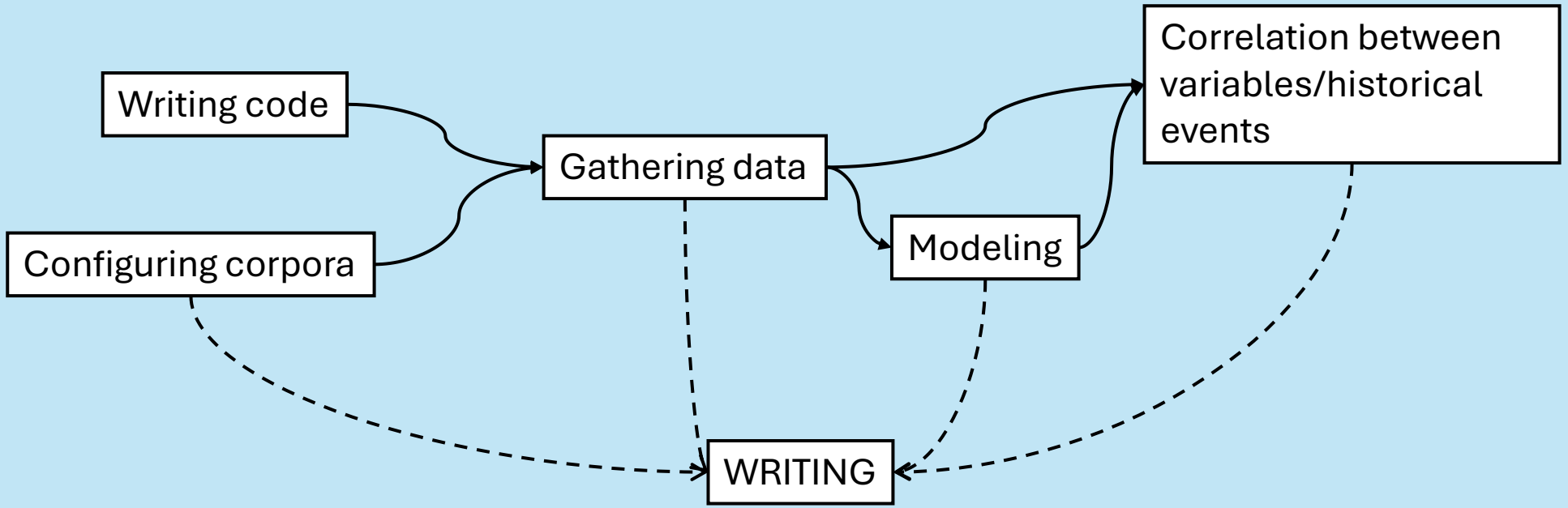
Comparing results for different significant variables

- Different variables will behave differently- which ones behave similarly? (same changes at the same times)
- Do mid-simulation changes correlate between variables? Do they correlate with moments of dramatic sociopolitical change?
  - This is both a question for the empirical work and the modeling
- Translating to the ‘properties’ of variables, e.g. hypernymization

5. The plan

“Work I will do”

What work needs to be done?



Schedule and chapter outline

- Summer 2025:
  - write Python scripts to efficiently gather distributional data
  - prep Mvskoke Creek material
  - Formatting Koasati texts from Kimball (2010)
- Fall 2025 and Spring 2026:
  - keep working Mvskoke material
  - start gathering data that can be identified automatically
  - manually tagging corpora for the variables that cannot be automated
  - writing background (Chapters 3 and 4, draft of Chapter 1)
- Summer 2026:
  - continue gathering data
  - finish historical profile/chapters
- Fall 2026 and Spring 2027:
  - (finish gathering data if necessary)
  - modeling
  - data analysis, writing Chapters 5-8
  - writing the historical relation chapters (Chapters 10, 11)

Chapter 1- Introduction
Chapter 2- History of contact and sociopolitical change in the Southeast
Chapter 3- Background [theoretical lit review]
Chapter 4- Measuring granular change [build on the prospectus]
Chapter 5- Trends in clausal morphology
Chapter 6- Trends in the range of syntactic structures
Chapter 7- Trends pertaining to hypernymy
Chapter 8- Trends in borrowing [build on QP2]
Chapter 9- More general conclusions from
Chapter 10- Summary of genre and social class localization
Chapter 11- Summary of localization within the lexicon
Chapter 12- Conclusion

References

- Bowern, Claire. (2012). *A Grammar of Bardi*. De Gruyter Mouton.
- Bromham, Lindell, Russell Dinnage, Hedvig Skirgard, Andrew Ritchie, Marcel Cardillo, Felicity Meakins, Simon Greenhill, Xia Hua. (2022). "Global predictors of language endangerment and the future of linguistic diversity." *Nature Ecology and Evolution*. Vol. 6. Pp. 163-173
- Brown, Cecil. (1999). *Lexical Acculturation in Native American Languages*. Oxford University Press.
- Epps, Patience. (2009). "Loanwords in Hup, a Nadahup language of Amazonia." in *The loanword typology project and the world loanword database*. Eds. Martin Haspelmath and Uri Tadmor. De Gruyter Mouton. Pp. 992-1014.
- Galloway, Patricia. (2009). "Choctaws at the Border of the Shatter Zone: Spheres of Exchange and Spheres of Social Value." In *Mapping the Mississippian Shatter Zone: The Colonial Indian Slave Trade and Regional Instability in the American South*. Eds. Robbie Ethridge and Sheri M. Shuck-Hall. University of Nebraska. Pp. 333-364.
- Haas, Mary. (1944). "Men's and Women's Speech in Koasati." *Language*. Vol. 20:3. Pp. 142-149.
- Haas, Mary. (1945). "Dialects of the Muskogee Language." *International Journal of American Linguistics*. Vol. 11:2. Pp. 69-74.
- Katenkamp, Seth. (2025). "Lack of prestige in use of English borrowings in nineteenth century Choctaw society." Qualifying paper. Yale University.
- Kirby, Simon. (1998). "Language evolution without natural selection: From vocabulary to syntax in a population of learners." *Edinburgh Occasional Papers in Linguistics*.
- Krifka, Manfred. (2021). "Layers of assertive clauses: Propositions, judgements, commitments, acts." In P*ropositionale Argumente im Sprachvergleich: Theorie und Empirie./Propositional Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues.* Eds. J. M. Hartmann and A. Wöllstein. Gunter Narr. Pp. 1-41.
- Meakins, Felicity, Xia Hua, Cassandra Algy, Lindell Bromham. (2019). "Birth of a contact language did not favor simplification." *Language*. Vol. 95:2. Pp. 294-332.
- Nichols, Johanna. (2009). "Linguistic complexity: a comprehensive definition and survey." *Language Complexity as an Evolving Variable*. Eds. Geoffrey Sampson, David Gil, Peter Trudgill. Oxford University Press. Pp. 110-126

Thank you to Claire, to the folks who have agreed to be on my committee (Jason, Alejandra, and Jack), to Aaron and Jonah for building the Historical Choctaw Corpus with me, and to Finn and Mike for teaching me a lot about math in the past few weeks!