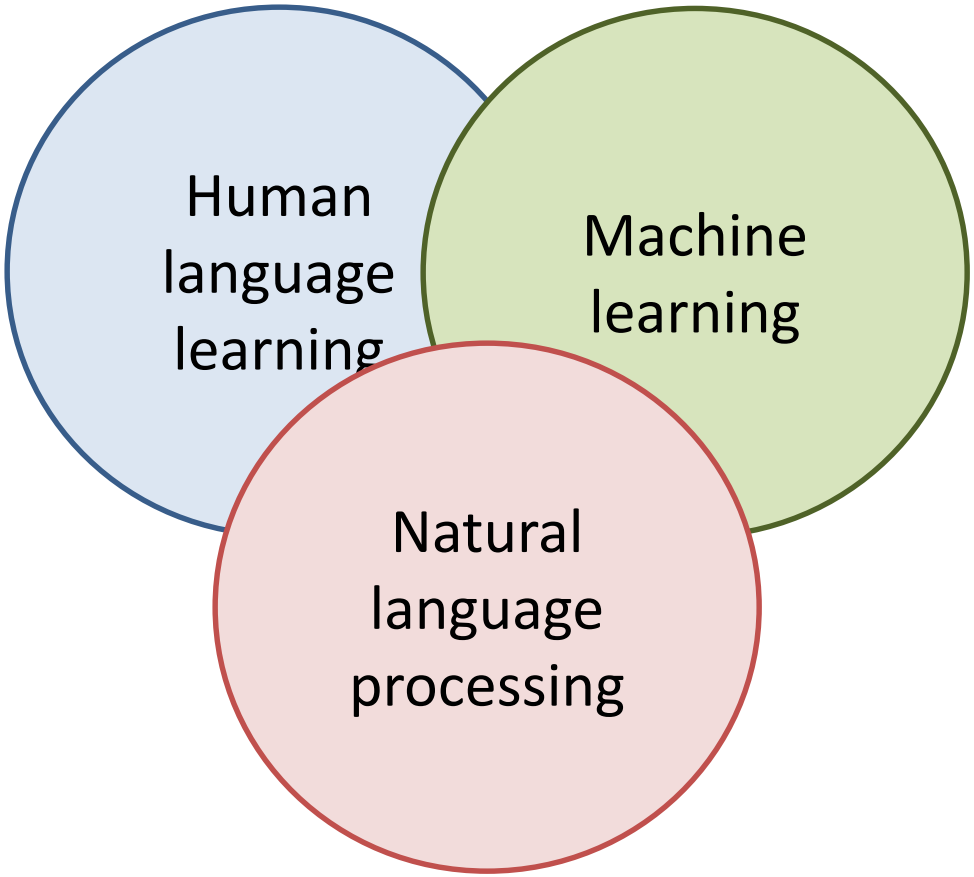
Models for language acquisition and change over time

Sharon Goldwater

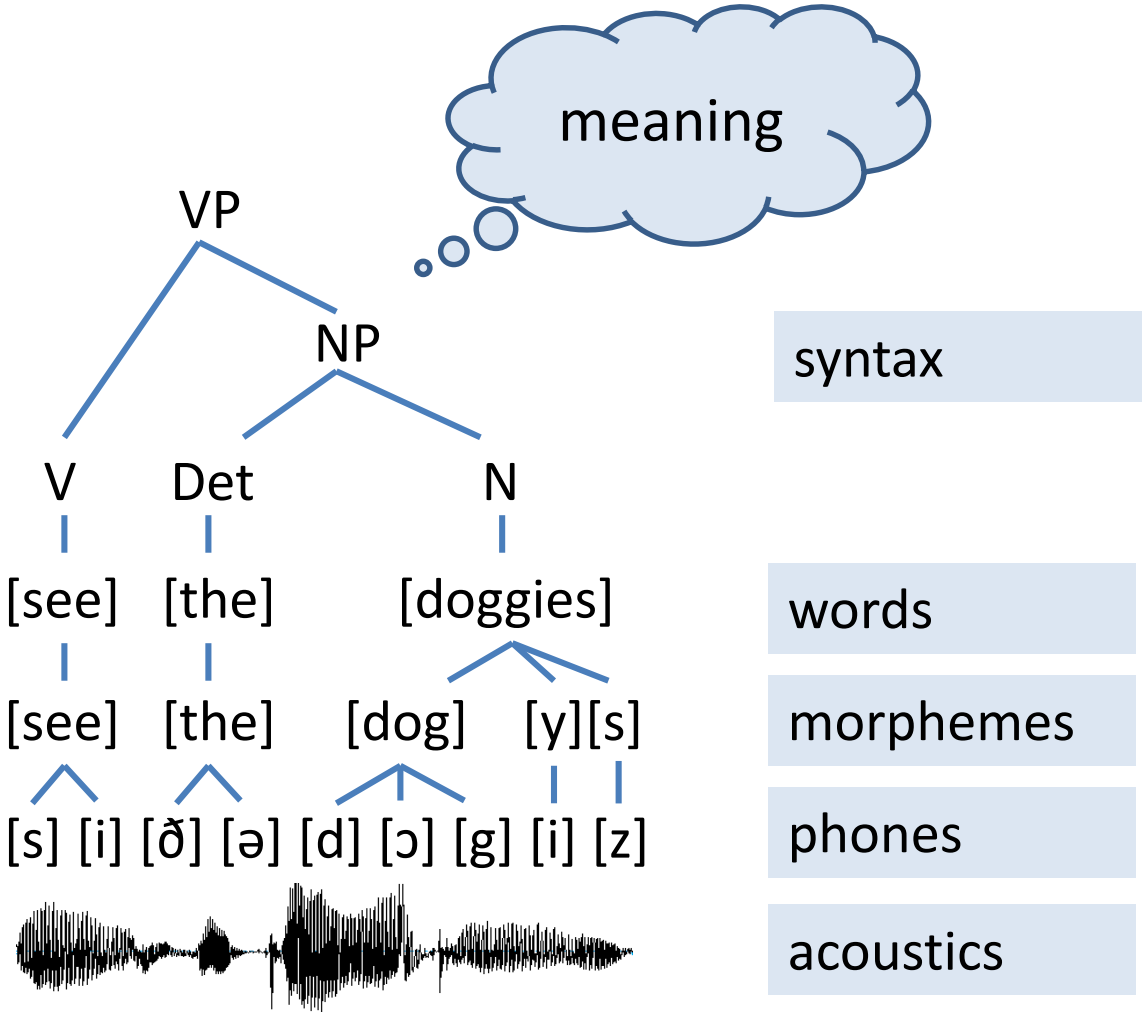
ilcc | Institute for Language, Cognition and Computation

THE UNIVERSITY of EDINBURGH
informatics

How can a computational system (whether human or machine) learn linguistic structure from linguistic data?



Linguistic structure

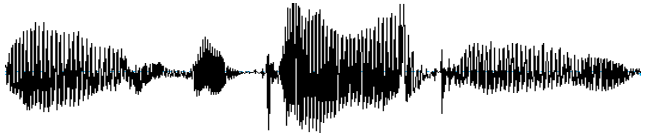


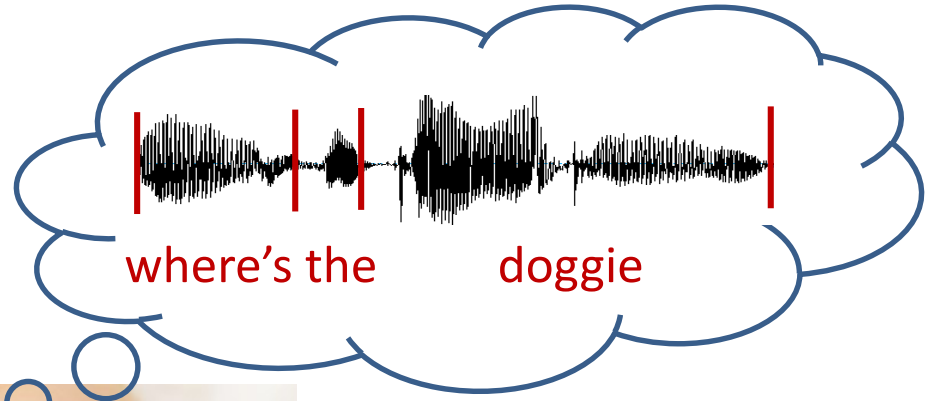
Linguistic data

- Raw acoustics or text, without annotations
- i.e., unsupervised
  - Like kids
  - Language processing for new languages
  - Useful ML models

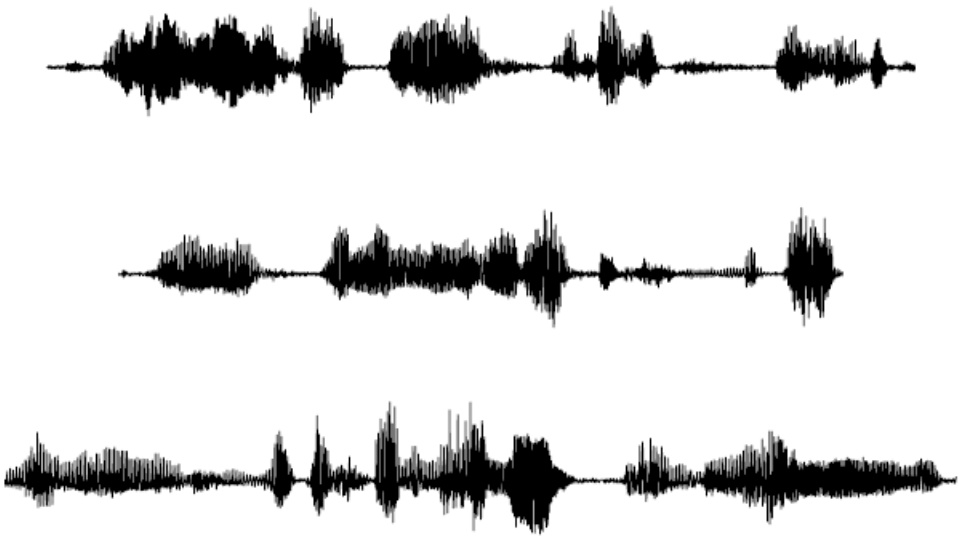
Ex 1: Word segmentation

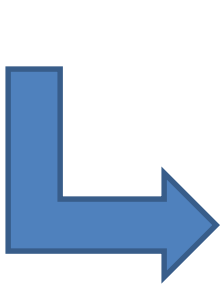
Current student: Herman Kamper

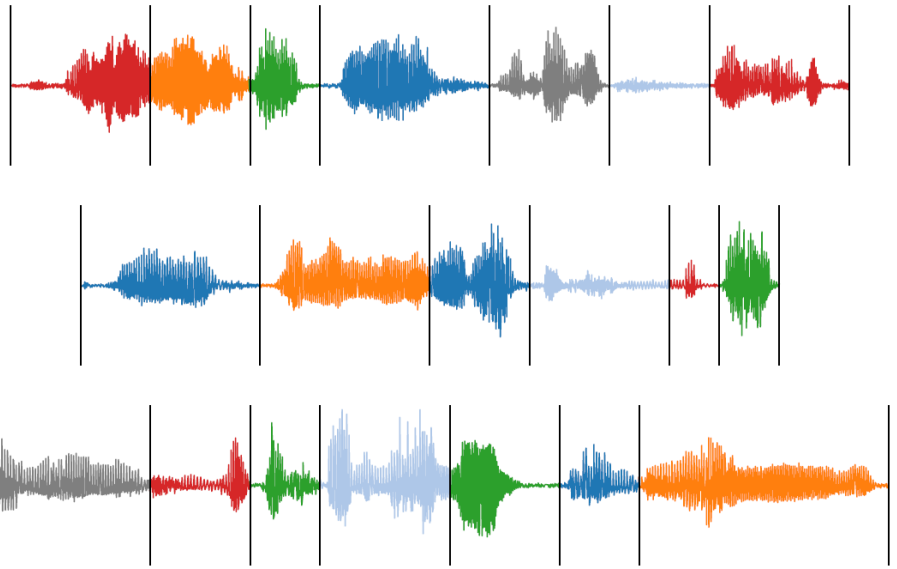




Segmentation and clustering







Early results

- Small-vocabulary corpus (TIDIGITS):

- Example output cluster:

Scaling up – what are the issues?

- We don’t know how many clusters.
  - Use nonparametric Bayesian models

Scaling up – what are the issues?

- We don’t know how many clusters.
  - Use nonparametric Bayesian models
- Difficult to cluster variable-length sequences.
  - Use fixed-dimensional representations
  - Project: Improve accuracy and efficiency (ANNs?)

Scaling up – what are the issues?

- We don't know how many clusters.
  - Use nonparametric Bayesian models
- Difficult to cluster variable-length sequences.
  - Use fixed-dimensional representations
  - Project: Improve accuracy and efficiency (ANNs?)
- Noise and irrelevant variability in speech.
  - Project: Learn better low-level representations (again, neural network methods?)

Cognitive science aspects

- What are infant's word representations like?
  - E.g., whole-word representations or phonetic sub-units? Proposals but no implementations.
  - Project: Test some claims from literature using whole-word representations from our model.
  - Project: Consider how to extend the model to learn sub-word phonetic units.

Cognitive science aspects

- What are infant's word representations like?
  - E.g., whole-word representations or phonetic sub-units? Proposals but no implementations.
  - Project: Test some claims from literature using whole-word representations from our model.
  - Project: Consider how to extend the model to learn sub-word phonetic units.
- How does non-linguistic context help?
  - Project: Extend the model to incorporate this type of information.

Ex 2: adult language adaptation

Current student: Philippa Shoemark

Ex 2: adult language adaptation

Current student: Philippa Shoemark

- Conversational partners align to each other:
  - Adapt speaking rate, use of vocabulary
  - But also deeper aspects like syntax

I gave the book to Joe	vs.	*I gave Joe the book*
I bought a cake for Mary	vs.	*I bought Mary a cake*

Ex 2: adult language adaptation

Current student: Philippa Shoemark

- Conversational partners align to each other:
  - Adapt speaking rate, use of vocabulary
  - But also deeper aspects like syntax
- And, languages change over time:
  - Again, both vocabulary and syntax

Because NOUN:
I didn't get much done today because internet.

Ex 2: adult language adaptation

Current student: Philippa Shoemark

- Conversational partners align to each other:
  - Adapt speaking rate, use of vocabulary
  - But also deeper aspects like syntax
- And, languages change over time:
  - Again, both vocabulary and syntax
- How do these processes relate to each other?
  - Use social media text to investigate.

More specific questions

- Which individuals in a social network are most responsible for spread of language change?
  - Those who align to others?
  - Those to whom others align?
  - Those who are more central to the social network?
- Data science methods can unite previously distinct areas of study.
  - Data dump from Reddit (and/or others)
  - Analysis tools from NLP and network science

Conclusion

- Lots of interesting work in this space, for lots of different backgrounds!