Take a Look!: Investigating the Relative Contributions of Children's Books and Child-Directed Speech

Joseph Denby

Computational Content Analysis

March 15, 2018

Setup / Objective

- Caregiver speech and picture books are two prominent sources of linguistic input for children
- Research has historically neglected the latter
- Important to assess their relative (unique?) contributions

Montag et al. (2015), 2

What language-learning data might early picture books provide that everyday conversations do not?

Corpora – LDP

- UChicago-based initiative to document parent-child interactions with a socioeconomically-diverse sample
- Ecological check-ins between ages 14 - 58 months at four month intervals
- Analysis draws from transcripts of caregiver speech

Corpora – Picture Books

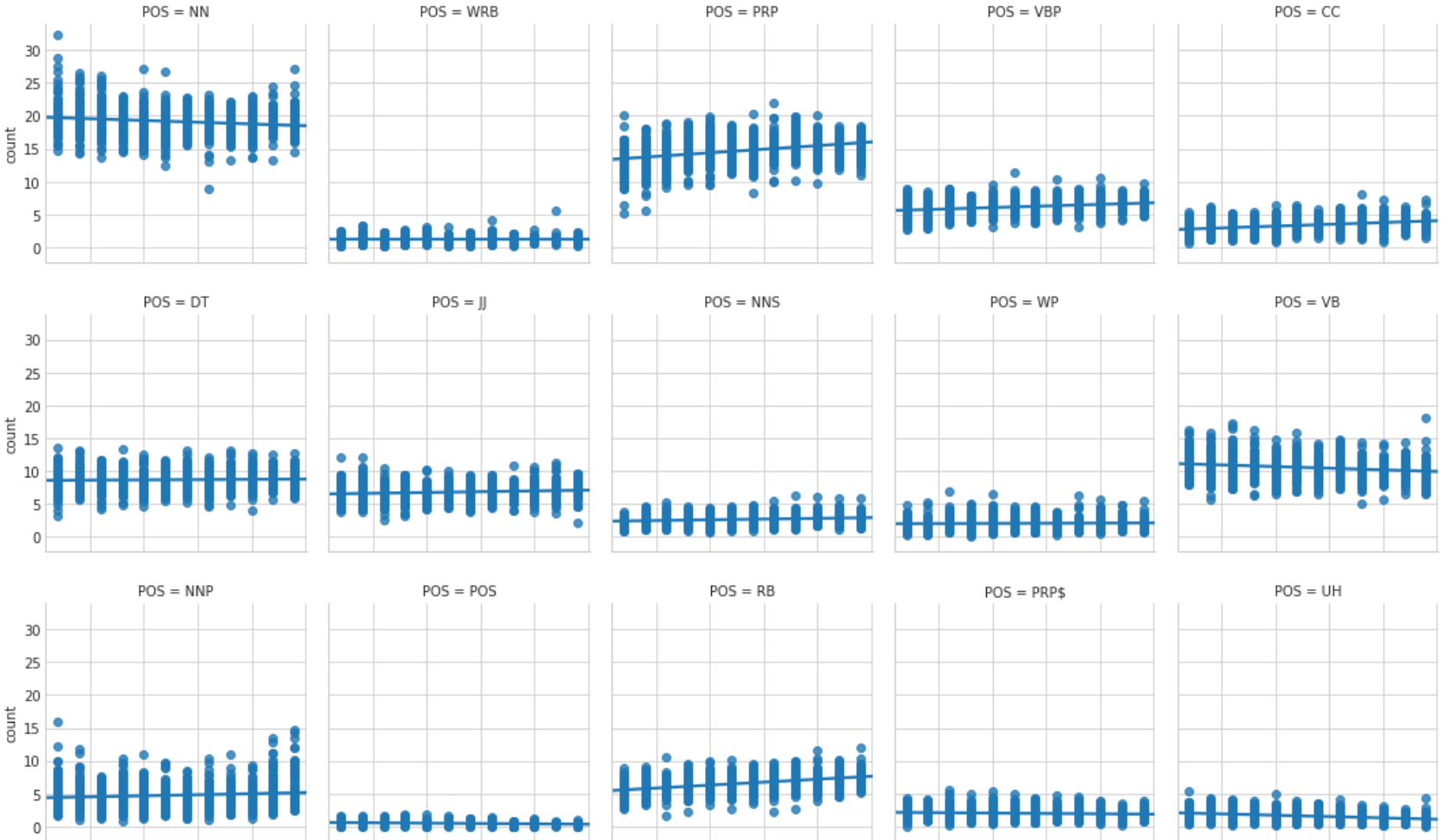
Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. Psychological Science, 26, 1489-1496.

Alexander and the Terrible, Horrible, No Good, Very Bad Day by Judith Viorst
Angelina Ice Skates by Katharine Holabird
Are You My Mother? by P. D. Eastman
Arnie the Doughnut by Laurie Keller
Arthur Writes a Story by Marc Brown
A Bad Case of Stripes by David Shannon
Bark, George by Jules Feiffer
Bear Wants More by Karma Wilson
The Berenstain Bears and the Green-Eyed Monster by Stan Berenstain and Jan Berenstain
The Berenstain Bears Forget Their Manners by Stan Berenstain and Jan Berenstain
Blueberries for Sal by Robert McCloskey
Bread and Jam for Frances by Russell Hoban
Brown Bear, Brown Bear, What Do You See? by Bill Martin, Jr.
Bunny Party by Rosemary Wells
Caps for Sale by Esphyr Slobodkina
The Carrot Seed by Ruth Krauss
The Cat in the Hat by Dr. Seuss
Charlie and the New Baby by Ree Drummond
Chicka Chicka 1-2-3 by Bill Martin, Jr., Michael Sampson, and Lois Ehlert
Chicka Chicka Boom Boom by Bill Martin, Jr., and John Archambault
Chrysanthemum by Kevin Henkes
How Do Dinosaurs Say Good Night? by Jane Yolen and Mark Teague
How to Train a Train by Jason Carter Eaton
If You Give a Moose a Muffin by Laura Joffe Numeroff
If You Give a Mouse a Cookie by Laura Joffe Numeroff
I'm a Big Sister by Joanna Cole
The Keeping Quilt by Patricia Polacco
Knuffle Bunny by Mo Willems
Ladybug Girl at the Beach by David Soman and Jacky Davis
Lilly's Purple Plastic Purse by Kevin Henkes
Little Blue Truck Leads the Way by Alice Schertle
The Little Engine That Could by Watty Piper
The Little House by Virginia Lee Burton
Llama Llama Home With Mama by Anna Dewdney
Llama Llama Red Pajama by Anna Dewdney
The Lorax by Dr. Seuss
Love You Forever by Sheila McGraw
Madeline by Ludwig Bemelmans
Maisy Goes Camping by Lucy Cousins
Maisy Goes to the Library by Lucy Cousins
Make Way for Ducklings by Robert McCloskey
Mike Mulligan and His Steam Shovel by Virginia Lee Burton
Miss Rumphius by Barbara Cooney
The Napping House by Audrey Wood
No, David! by David Shannon
Oh, the Places You'll Go by Dr. Seuss

Outline & Sampling

1. Compare POS usage between the two corpora
   - Randomly sample 150-word contiguous windows from books and conversations
2. Analyze and compare sentence complexity within and between corpora
   - Randomly sample 90-sentence contiguous windows from conversations (approximate mean book length)

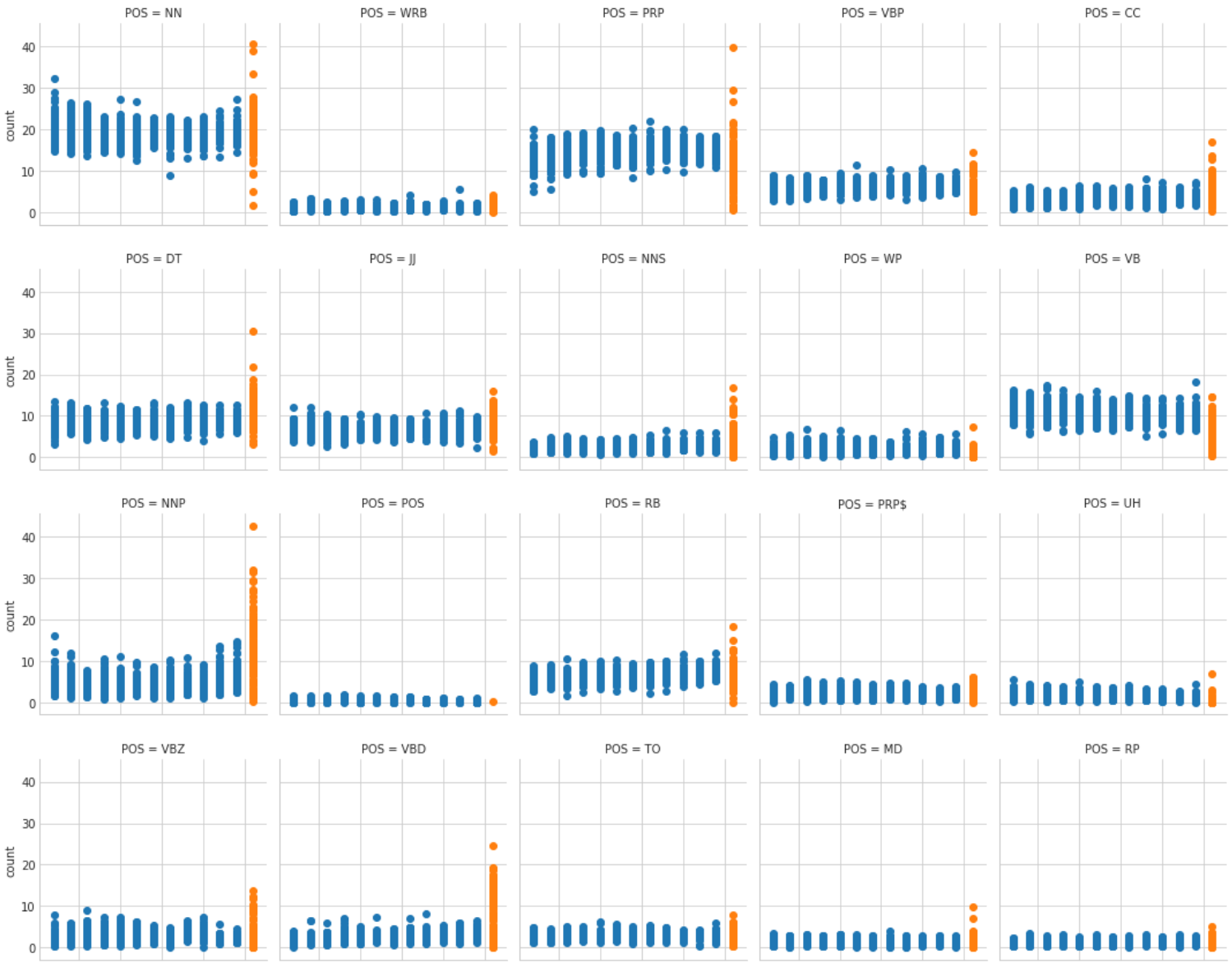
Analysis – Chat POS (14-58mth, 4mth interval)



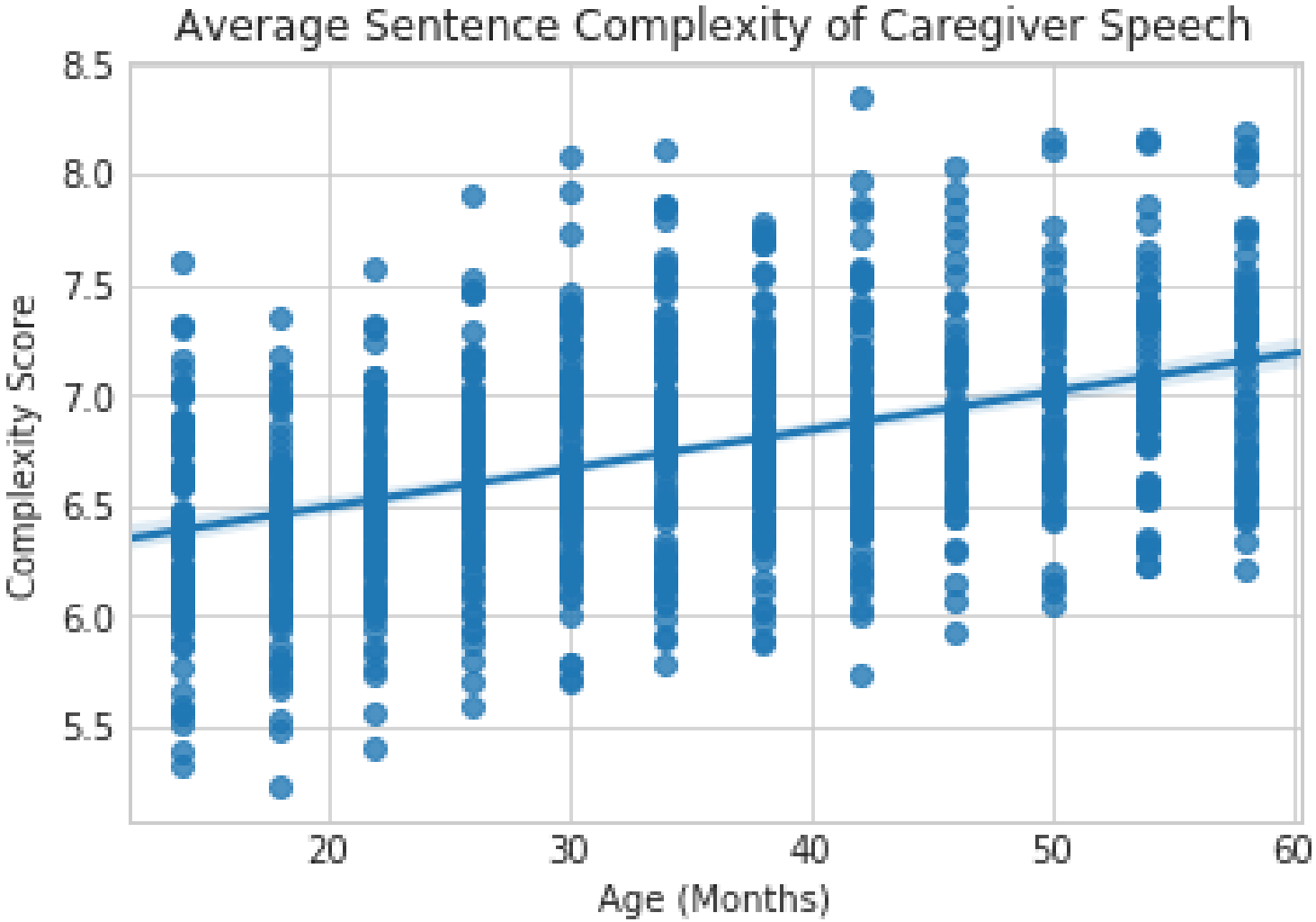
Analysis – Chat&Book POS (14-58mth, 4mth interval)

Analysis – Chat&Book POS (14-58mth, 4mth interval)

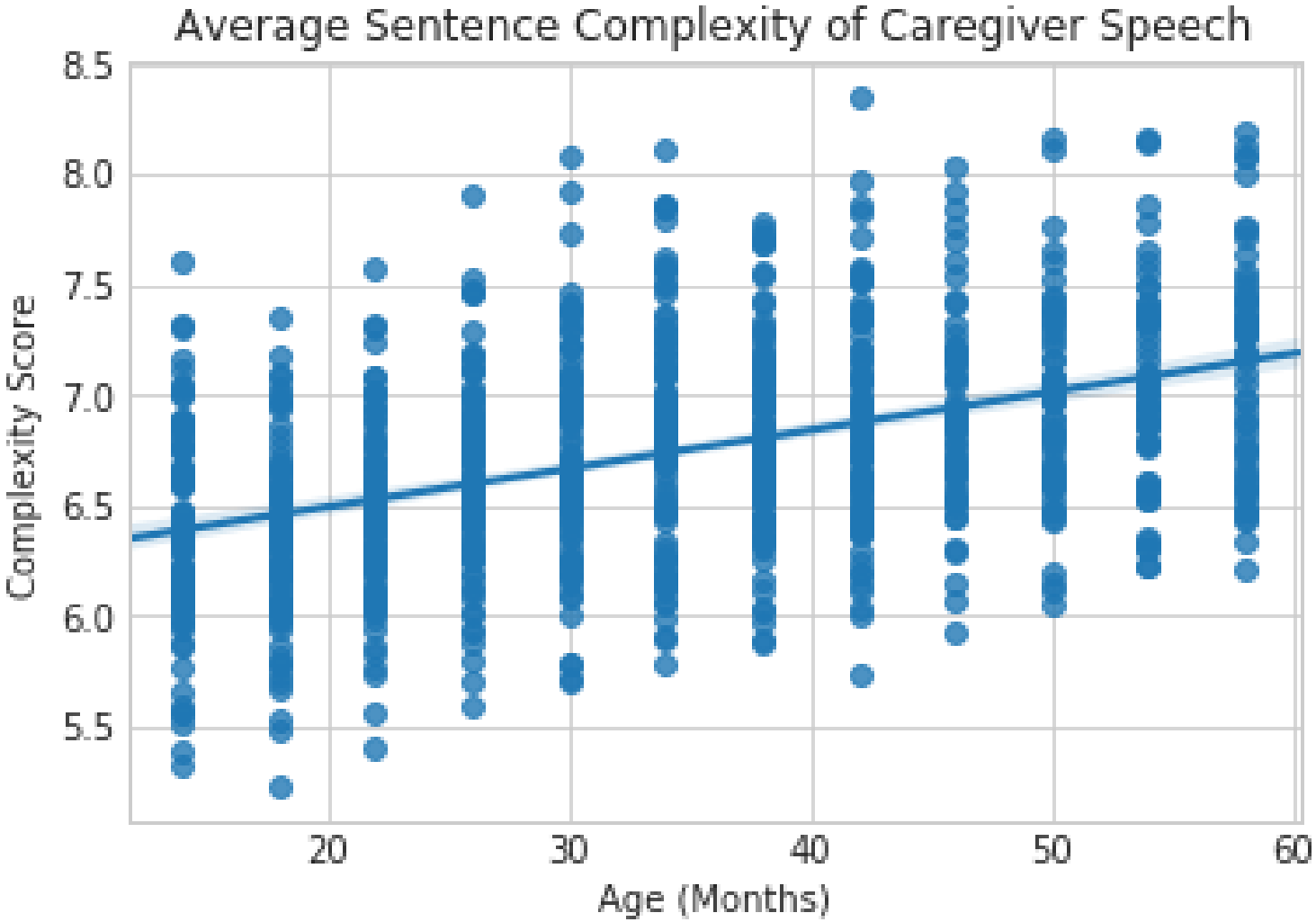
Analysis – Chat&Book POS (14-58mth, 4mth interval)



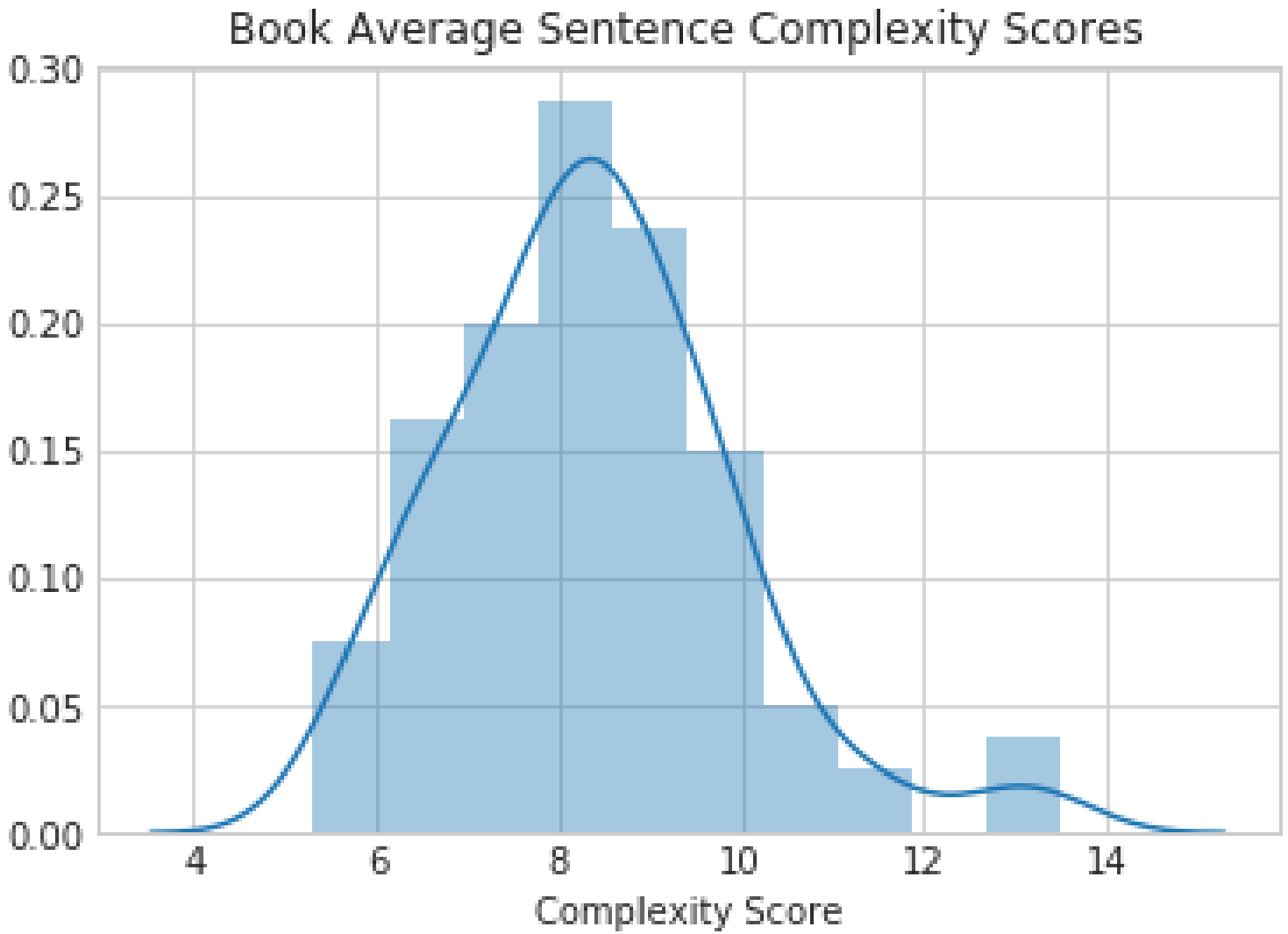
Analysis – Chat Complexity



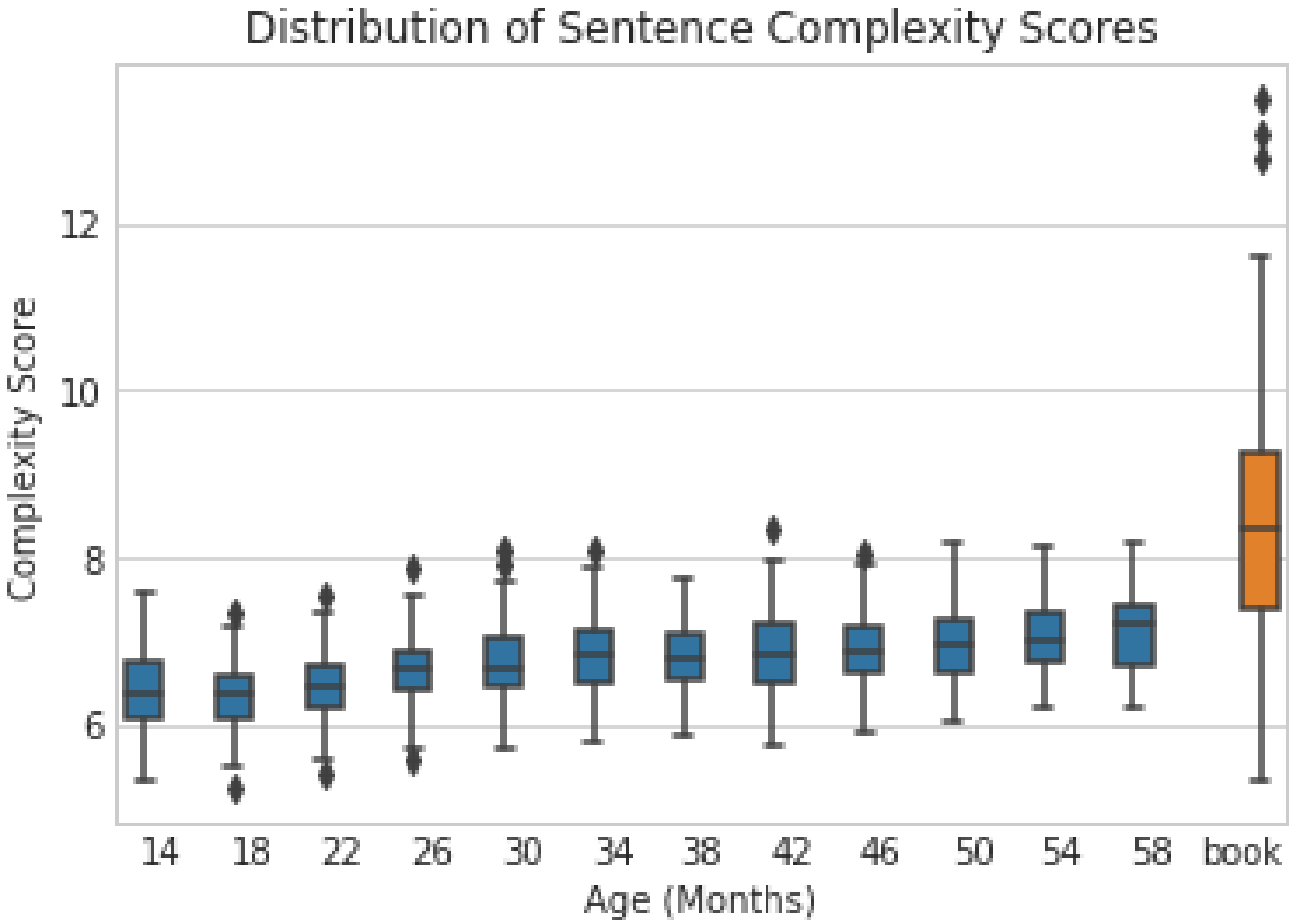
Analysis – Chat Complexity



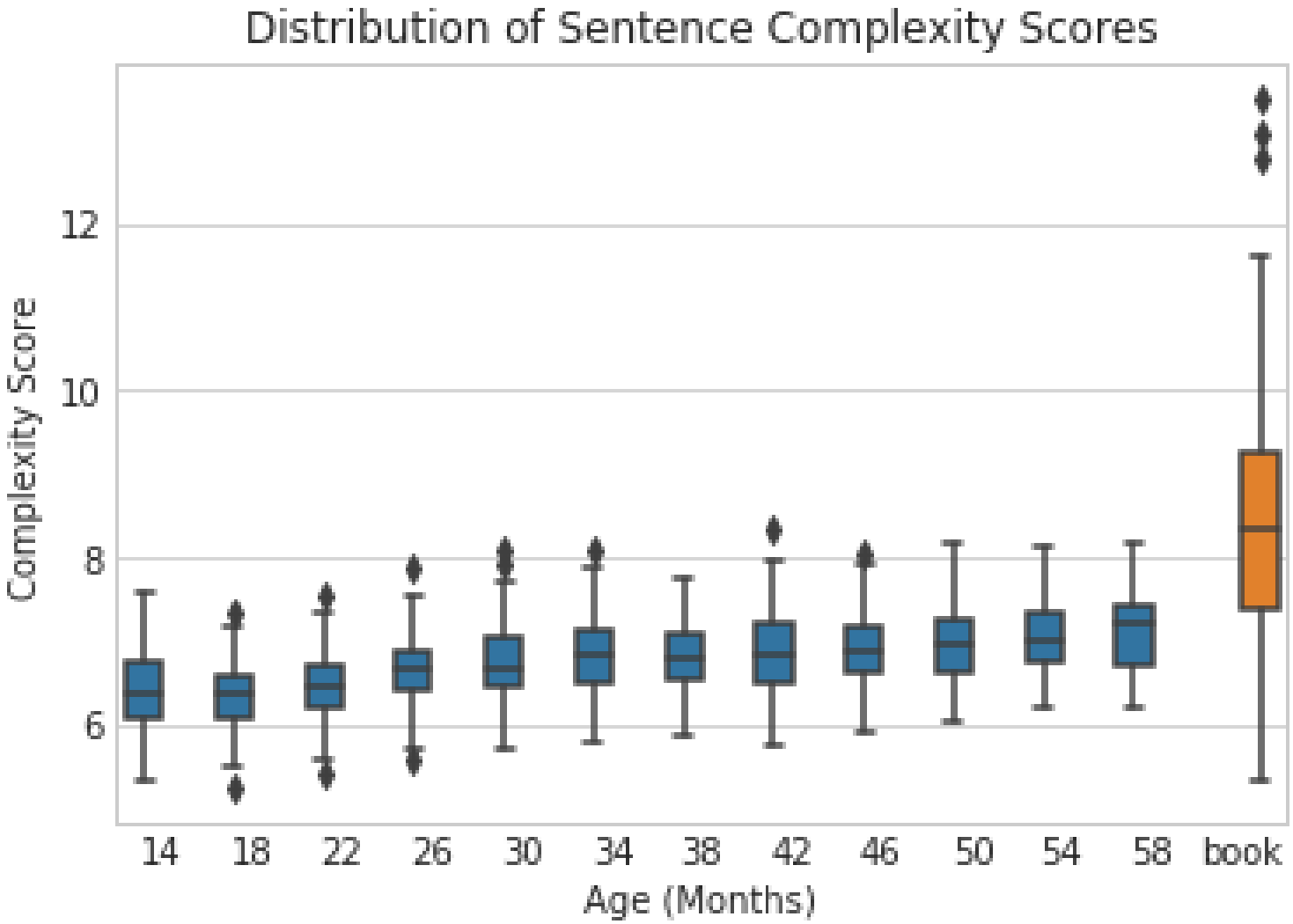
Analysis – Book Complexity



Analysis – Chat&Book Complexity



Analysis – Chat&Book Complexity



Investigation

Linguistic Coordination

Given speakers a, b, with utterances $u_1 : a, u_2 : b$, and linguistic category m :

$$C^m(b, a) = P(\varepsilon^m_{u_2 \hookrightarrow u_1} | \varepsilon^m_{u_1}) - P(\varepsilon^m_{u_2 \hookrightarrow u_1})$$

Simply, the likelihood that a speaker uses a linguistic category given that their interlocutor used that category, versus their baseline usage.

Power Claims (again)

For a target group G and arbitrary group U:

- **Powerful Speaker(s)**: $C(G^{high}, U) < C(G^{low}, U)$
- **Powerful Listener(s)**: $C(U, G^{high}) > C(U, G^{low})$

Wikipedia Results 1

Wikipedia Results 2

Supreme Court Results 1

Supreme Court Results 2

SVM Prediction

Coordination features were only features that performed significantly above chance for cross-domain prediction.

Brief Summary

- The two power claims were supported:
  - Powerful speakers are more aligned-to
  - Powerful listeners do less aligning
- Changes in power produce different alignment behavior (after controlling for baselines)

Potential Questions

What about other networks?

Do dispersed social networks (e.g., Facebook, Twitter) show similar linguistic markers of structural power?

Twitter Alignment

Power by Followers

Power by Verification

Doyle, G., Yurovsky, D., Frank, M. C. (2016). A robust framework for estimating linguistic alignment in Twitter conversations. Proceedings of the 25th International Conference on World Wide Web, 637648.

Potential Questions

What constitutes a power relation?

Danescu-Niculescu-Mizil et al., 1

Our focus is on domains in which groups engage in goal-oriented discussions...An important characteristic of such discussions is that the participants are invested in the issues at hand, so that their dialogs are not simply idle chat.

- What is 'idle chat?' Why is it disqualified?

Further Work

To evaluate whether power-based linguistic coordination, evaluate casual conversations *without* explicit goals/tasks/choices.

E.g., off-the-record conversations, water cooler talk, etc.

Potential Questions

What other purpose(s) does linguistic coordination serve?
Is it always characterizable as evocative of power structure?

Linguistic Alignment in Parent-Child Interactions

callab.uchicago.edu

Linguistic Alignment in Parent-Child Interactions

What constitutes a power structure?

Is it fair to characterize the parent-child dynamic as 'dependent?'