The (Un)Surprising Kindergarten Path

Zoe Ovans (zovans@umd.edu)
Program in Neuroscience and Cognitive Science &
Department of Hearing and Speech Sciences, 0100 Samuel J. LeFrak Hall
College Park, MD 20742 USA

Yi Ting Huang (ythuang1@umd.edu)
Department of Hearing and Speech Sciences, 0100 Samuel J. LeFrak Hall
College Park, MD 20742 USA

Naomi H. Feldman (nhf@umd.edu)
Department of Linguistics and UMIACS, 1401 Marie Mount Hall
College Park, MD 20742 USA

Abstract

During sentence comprehension, listeners form expectations about likely structures before they have reached the end of a sentence. Children are more likely than adults to ignore late-arriving evidence when it contradicts their initial parse. While this difference is often ascribed to developmental changes in executive function, this paper investigates whether statistical properties of child-directed speech could be responsible for children's failure to revise temporarily ambiguous sentences. We examined well-studied garden-path sentences and calculated surprisal values derived from adult and child-directed corpora at each word. For adult corpora, surprisal was highest where the sentence structure was disambiguated. For child corpora, however, values at the disambiguating region were low relative to other words in the sentence. This suggests that for children, the disambiguating words may be statistically weak cues to ruling out their original parse, and that in principle, the statistics of child-directed speech could contribute to children's difficulty with garden-path sentences.

Keywords: Language; language development; corpus studies; kindergarten-path; surprisal

1. Introduction

1.1 The Kindergarten-path effect

During language comprehension, listeners interpret sentences incrementally, integrating each new word in real time as a sentence unfolds (Frazier & Rayner, 1982, *inter alia*). Incremental processing can make comprehension difficult: since listeners are constantly guessing the structure of the full sentence before they've heard every word, they may be surprised when late-arriving words are inconsistent with initial parses (Rayner, Carlson & Frazier, 1983; Pickering & Traxler 1998; Sturt, Pickering & Crocker, 1999). These dynamics are visible when listeners encounter temporarily ambiguous or "garden-path" sentences. For example, in (1), listeners are initially lulled into assuming the wrong attachment for the first prepositional phrase (PP1: "on the napkin").

(1) Put the frog on the napkin into the box

At first, PP1 seems like it should attach to the verb phrase (VP), and offer a location to the "putting" event (as in "Put it on the napkin."). However, the subsequent arrival of PP2 ("into the box") indicates that this interpretation was incorrect, and that PP1 instead attaches to the first noun phrase (NP1) (as in "Put the frog *that is* on the napkin…"). Given a display like Figure 1, adults often initially look to the incorrect goal (the empty napkin) at "napkin" before looking to the correct one (the box) at "into" (Tanenhaus et al., 1995; Trueswell et al., 1999). When asked to act out the action after hearing the sentence, adults have little difficulty putting the frog directly into the box. This indicates that while they temporarily consider the incorrect VP-attachment parse, they quickly revise their interpretation in favor of the ultimately correct NP-attachment interpretation after disambiguation. Psycholinguistic models of surprisal have been used to quantify adults' temporary confusion and the extent to which late-arriving words rule out earlier parses by correctly predicting that the location of increased processing costs will be at the start of PP2, where the sentence is disambiguated (Levy et al., 2008).

Children have greater difficulty than adults in recovering from these sentences. In a seminal study, Trueswell and colleagues examined 5 year-olds' interpretations of sentences like (1) using a visual-world eye-tracking paradigm. Children and adults listened to these sentences while viewing a visual display like Figure 1. After PP1, they looked between the napkin-less frog and the empty napkin, indicating a VP-attachment interpretation (i.e. "Put it on the napkin"). After PP2, adults quickly shifted their gaze to the box, suggesting

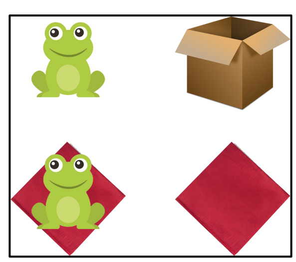

Figure 1: Schematic of a display for Trueswell et al., (1999)

revision in favor of an NP-attachment interpretation. In contrast to adults, after PP2, children were significantly slower to look away from the incorrect goal, indicating that they had more difficulty overriding their initial incorrect parse. After the sentences, children also enacted the events incorrectly, often by putting the animal onto the incorrect goal before the correct one (e.g. hopping a frog to the empty napkin, and then to the box). This phenomenon is known as the "Kindergarten-path" effect (Trueswell et al., 1999; Anderson et al., 2011; Choi & Trueswell, 2010; Hurewitz et al., 2000; Weighall, 2008).

1.2 Accounting for the Kindergarten-path effect

The standard explanation for children's difficulty with garden-path sentences is that children, like adults, parse sentences incrementally, and subsequently must revise incorrect parsing decisions in the face of late-arriving conflicts. But unlike adults, their underdeveloped cognitive control system prevents them from inhibiting the initial parse they built (Novick et al., 2005; Woodard et al., 2016; Choi & Trueswell, 2010; Mazuka et al., 2009). Specifically, under the cognitive-control account, children's difficulty stems from errors in mediation between two conflicting parses. Systems for generating parses and detecting errors may be identical across development, but once children realize that the particular sentence representation they've built is incorrect, errors arise due to their failure to inhibit the initial parse. Under a strict interpretation of this account, the Kindergarten-path effect arises solely due to differences in adults' and children's nonlinguistic cognitive abilities. It is not attributable to other factors, such as the linguistic expectations children and adults generate based on what they have inferred about their language from their input.

However, evidence is mixed for the claim that cognitive control differences *fully* account for children's garden-path errors. For one, performance on non-linguistic tests of cognitive control – including the Day-Night task (Gerstadt et al., 1994), Dimensional Change Card Sort (Jacques & Zelazo, 2001), and the Flanker task (Eriksen & Eriksen, 1974) – does not consistently correlate with success at parsing garden-path sentences (Woodard et al., 2016; Qi et al., 2020). Second, when children's overall language abilities are equated based on vocabulary size, individual variation in cognitive-control skills has no discernable impact on parsing garden-path sentences (Huang et al., 2017). While inhibitory processing skills likely play a role in successfully parsing temporary ambiguities, these results suggest that other developmental factors beyond domain-general cognitive control maturation may also be needed to explain children's difficulties.

In this paper, we investigate a new hypothesis: that children's difficulty with garden-path sentences may be in part due to differences between their input and that of adults. Parental input often reflects the communicative goals of interacting with less linguistically and cognitively developed interlocutors (Newport, Gleitman, & Gleitman, 1977; Lieven, 1984; Richards, 1994). Child-directed speech tends to use short sentences (Snow & Ferguson, 1977), and refer to events in the here-and-now through imperatives (Warren-Leubecker et al., 1984) and pronouns (Tardif, Shatz, & Naigles, 1997; Laakso & Smith, 2007). The strict cognitive-control account implicitly assumes that these differences in input do not play a role in accounting for the Kindergarten-path effect, but this has not been directly investigated.

To do so, we quantitatively measure differences between children's and adults' input using surprisal, a well-known psycholinguistic measure of the probability of words given their sentential context. This metric quantifies the extent to which developmental differences in input impact parsing by taking into account the structural and lexical frequencies present in adult and child-directed corpora. Surprisal offers a way to measure the extent to which listeners' prior experience with their input might affect their expectations about each new word as they're parsing a sentence. We use surprisal to test for differences between the expectations generated by adult and child-directed input, for a set of relevant garden-path sentences that have been used to measure the Kindergarten-path. Importantly, the errors that children make in processing "Put" sentences are likely due to the subcategorization requirements of the verb "put" (namely, that it requires a location). Thus, prior to the surprisal calculation, we semi-lexicalized our corpora by tagging the VP nodes with information about the head verb.

To preview our results, we find an expected pattern for adults: highest surprisal values fall on the words that disambiguate the sentences. For children, however, we find that the disambiguating regions have relatively lower surprisal values, relative to other words in the sentences. This indicates that variation between adult and child-directed speech could lead to differences in how adults and children process these sentences in real time. Thus, children's trouble with Kindergarten-path sentences relative to adults may stem not just from differences in cognitive control, but also from differences in what they have learned about how verbs predict subsequent words in their language, based on their input.

2. Interpreting the input signal with Surprisal

Hale (2001) proposed that a word's reading time is proportional to its surprisal, computed as the natural log of its conditional probability of appearing in an utterance. In other words:

$$\text{Reading time} = -\alpha \log p(w_i | w_1 \dots w_{i-1})$$

Where α is a proportionality constant, w_1 is the first word in the sentence and w_i is the word whose surprisal value is being calculated.

This measure has been used frequently in previous work as an index of processing difficulty, with more surprisal predicted at words where highly probable parses are ruled out. Traditional accounts of why surprisal values track reading-time performance assume that there is a processing cost incurred when listeners encounter input that is inconsistent with the sentence structure they'd been building. This cost is incurred because of a mismatch between a new word that's encountered, and the expectations listeners had built from their prior input. For adults, we expected to find high surprisal at the disambiguating word of temporarily ambiguous sentences.

Importantly, under a strict cognitive-control account, children and adults are both good at *detecting* when late-arriving words rule out an initial parse, but immature cognitive-control impairs correct selection of competing alternatives. Translating this to a surprisal model, the cognitive-control account predicts that the disambiguating region of the sentence should be unexpected to both adults and children, and children's difficulty lies only in domain-general conflict resolution. Alternatively, finding a different pattern at the point of disambiguation for adult and child data may indicate that the statistics of child-directed speech are different from adult speech in a way that matters for these temporarily ambiguous structures.

3. Methods

We trained language models using parsed corpora of adult- and child-directed speech. The grammars that were learned from each of these corpora approximate the structural and lexical frequencies present for adults and children, respectively. These grammars were then used to calculate surprisal values at each word. The surprisal values provide a composite measure of how unexpected each word is: they take into account both whether high probability parses are ruled out at that word, and the rarity of the specific lexical item.

For adult-directed speech, we expect to replicate previous work and find that the region in the sentence with the highest surprisal value will be the point of disambiguation. We are interested in whether child-directed speech shows a different pattern.

3.1 Training and evaluation datasets

To determine the surprisal values at each word for adults, we used a sample of 118,000 sentences from the Switchboard corpus (Godfrey et al., 1992). For child-directed speech, we used the Pearl & Sprouse parsed corpus of child-directed speech (Pearl & Sprouse, 2013). This corpus contains parsed versions of six corpora within the CHILDES database (MacWhinney, 2000). The data represent speech to children between the ages of 1 and 5 years, and contain over 150,000 total utterances, and is to date the largest parsed corpus of child-directed speech. We trained a separate model on each of these corpora, for a total of six language models trained on child-directed speech.

Our test set consisted of 24 sentences with PP-attachment ambiguity on which children have been shown to have difficulty reaching the correct final interpretation (Weighall, 2008, Appendix A items 1-24). All sentences were similar in structure to (1), and contained "put" as a matrix verb. Each sentence contained different nouns that referred to common animals, objects, and locations and were known to typically developing 5-year-olds (e.g. "Put the fish on the sponge on the plate"). Prepositions were either "in," "on," "behind," or "with."

3.2 Calculating surprisal values

Surprisal values at word *i* were calculated as

$$-\log_2\left[\frac{p(w_1 \dots w_i)}{p(w_1 \dots w_{i-1})}\right]$$

Where $p(w_1 \dots w_{i-1})$ is computed by summing the probabilities of all possible parses that are compatible with the first i-1 words of the sentence, given the grammar, and $p(w_1 \dots w_i)$ is computing by summing the probabilities of the subset of those parses that are also compatible with word i. Intuitively, lower-frequency words and words that rule out highly-probable parses generate higher surprisal values, since in these cases the prefix probability at word i will be much lower than the probability at word i-1. These probabilities were calculated using the EarleyX parser, a java implementation of Stolcke's probabilistic Earley parser (Stolcke, 1995; Earley, 1970). This parser provides a way of computing surprisal values from any imported phrase structure grammar, and provides several advantages over Stolcke's original algorithm (Luong et al., 2013). It is based on the prefix probability parser by Levy (2008), but uses an updated scaling method to reduce the time needed to parse long utterances.

We were concerned that surprisal values based on a probabilistic context-free grammar (PCFG) might not capture the desired effect, because the verb "put" is a determining factor in making the continuation "on the napkin" in (1) ambiguous. Structures that contain other verbs, such as "choose" (i.e., "Choose the frog on the napkin"), are not ambiguous. A typical PCFG does not take into account the fact that individual lexical items can influence which structures are possible, and so it might not accurately capture listeners' expectations that "put" should be followed by a location that describes the endpoint of the putting event. In order to ensure that our surprisal calculation was taking these probabilities into account, the parser was semi-lexicalized: each VP node was tagged with information about the verb head prior to calculating the prefix probabilities.

4. Results

Our adult and child corpora differ in many respects, including the lengths of utterances, the topics they contain, and the conversational setting (in-person versus over the phone). Because of this, we did not directly compare surprisal levels in the adult corpora to surprisal levels in the child-directed corpora. Instead, we compared the surprisal across the different words in the sentences within each corpus.

4.1 Adult-directed corpus results

Our first goal was to confirm that surprisal coincides with adults' slowdowns in processing sentences with PP-attachment ambiguity. We predicted that for adults, the

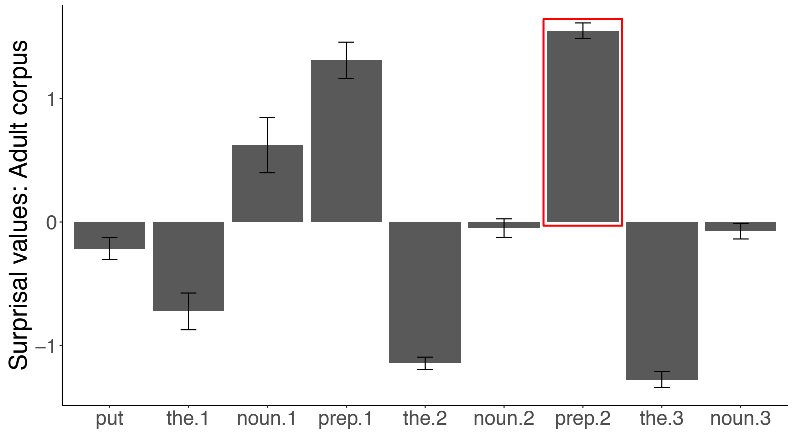

Figure 2: Z-scored surprisal values based on the adult Switchboard corpus for 24 garden-path sentences (from Weighall, 2008). Error bars represent standard error of the mean.

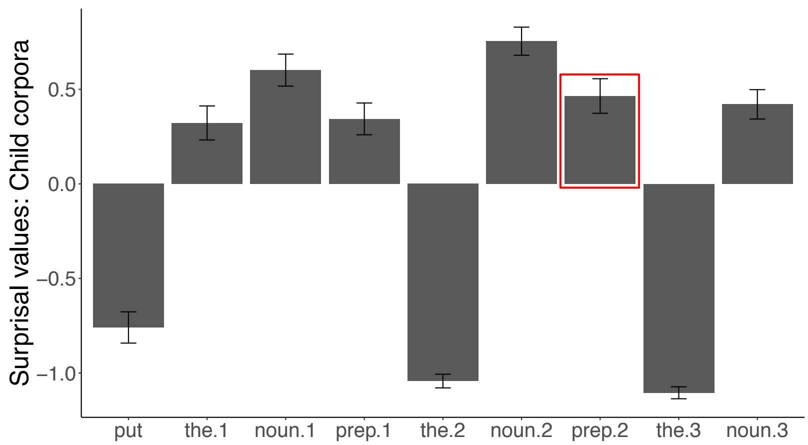

Figure 3: Z-scored surprisal values based on 6 child corpora for 24 garden-path sentences (from Weighall, 2008). Error bars represent standard error of the mean.

highest surprisal values in the sentence would be found at the start of PP2, where the sentence is disambiguated. Figure 2 shows normalized average surprisal values for adults for the 24 sentences with PP-attachment ambiguity (Weighall, 2008). These values were calculated by subtracting the mean value for each sentence from every word in the sentence, and dividing by the standard deviation of surprisal values in the sentence. Sensibly, the lowest surprisal values were found at definite determiners following the verbs and prepositions, since the parser takes lexical frequency into account. Importantly, the highest surprisal values for the adult corpus were indeed found at the disambiguating region (highlighted in the red box). This reaffirms what prior literature has found, that surprisal values track input-driven effects of expectations. Given that these values reflect the expected garden-path slow-downs, any differences between these results and those obtained from corpora of speech to children suggest that children have license to form different expectations based on the input they receive.

4.2 Child-directed corpus results

We next calculated surprisal values at each word based on a sample of speech to children. Figure 3 shows surprisal values for child-directed speech for the same 24 sentences as Figure 2, averaged across the six language models that were trained on child-directed speech. These values were calculated by subtracting the mean value for each sentence from every word in the sentence, for each corpus, and dividing by the standard deviation of surprisal values in the sentence. Generally, the surprisal results for the child corpora mirror those for adults, with a few notable exceptions. Similar to adults, lowest surprisal values occur at the definite determiners, suggesting that the frequency of these words is successfully taken into account. The exception is the first determiner, likely because children often hear “put” imperatives with pronouns (e.g., “put it down”) (Warren-Leubecker et al., 1984). However, unlike adults, surprisal at sentence initial “put” is relatively low for children, consistent with the observation that adults often use more imperatives when speaking to children than when speaking to other adults (Warren-Leubecker et al., 1984).

Most importantly for our study, surprisal values for children were not highest at the disambiguating regions of the temporarily ambiguous sentences (highlighted in the red box). Instead, for children, the initial nouns in the sentence consistently have surprisal values greater than the values at the point of disambiguation. This is likely in part due to the fact that our calculation of surprisal included a measure of lexical frequency, and for children, each noun token was relatively unexpected. Since the adult and child corpora we used contained utterances of different lengths and topics, we did not further analyze direct comparisons between the adult and child results at each word.

5. Discussion

The goal of this paper was to determine whether differences in adult and child-directed speech parallel developmental difficulties with garden-path sentences. We chose a set of consistent and paradigmatic garden-path sentence stimuli: all of our sentences contained PP-attachment ambiguity and have been shown to elicit larger errors for children than for adults during an eye-tracking task (Trueswell et al., 1999; Weighall, 2008). We used surprisal to measure how unexpected each word should be, based on the sentences found in adult and child corpora. We found that for adults, surprisal is highest on the words where a highly likely parse is ruled out. However, for child-directed speech, less surprisal is predicted at this disambiguating region than at other words in the sentence. This finding indicates that variability between what children and adults hear may contribute to children's garden path errors. While further work is needed to determine how such input differences might map onto the established real-time processing delays, the present results offer some suggestions.

Surprisal tracks the probability that a given word will occur next, and it is generally taken to be an index of the fact that a probable parse has been ruled out (Hale, 2001; Levy, 2008; Fine et al., 2013; Ferraro & Van Durme, 2016; Futrell et al., 2017). However, this account of surprisal is difficult to reconcile with the present data. Our analyses of child-directed corpora showed low surprisal at the disambiguating region. If lower surprisal values index relatively expected words, then we might conclude that children are already predicting NP-attachment (i.e., they are not surprised by it) and thus that they should have *less* difficulty than adults at reaching this interpretation. This is the opposite of what actually happens: this disambiguating region is where children display *more* difficulty than adults in reaching a final interpretation (Trueswell et al., 1999; Anderson et al., 2011; Choi & Trueswell, 2010; Hurewitz et al., 2000; Weighall, 2008).

Instead, our results are more consistent with an account that interprets high surprisal as reflecting a type of error signal. Under such an account, children would be receiving a weaker error signal than adults at the disambiguation point. If "into" in (1) is relatively expected compared to other words the child is hearing in the sentence, it may not be an effective signal that reparsing is needed at all. This view is relatively incompatible with previous characterizations of surprisal, which assume that difficulty arises when comprehenders must jump to a less expected parse or rebuild their parse tree from an earlier node. Instead, interpreting surprisal as an error signal assumes that difficulty can arise when this signal isn't encountered to begin with. We found that for child-directed speech, high surprisal is predicted at open-class words like nouns in these sentences. The signal that children get from the word that rules out a likely parse is then in competition with many other error signals in the sentence, and may be less salient. If they find the first nouns more unexpected than the point of disambiguation, they may fail to encounter the strong error signal that adults receive at the disambiguation point in the sentence. This is a different characterization of surprisal than is typically found in the sentence-processing literature, and the idea that low surprisal may reflect a failure to realize that a particular parse is incorrect would need to be reconciled with the fact that surprisal values are computed in the first place by summing across different possible parses of a sentence.

This is an explanation for the Kindergarten-path effect that has not been previously proposed in the literature on children's sentence processing. It suggests that Kindergarten-path effects occur because the error signals children receive are competing for their attention, and this added noise leads to a failure to detect the disambiguation. This assumes that the error signal children receive from encountering an infrequent noun is qualitatively similar to the error signal they receive when a likely parse is ruled out, and that this causes the point of garden-path disambiguation to be a weak cue to adopt a less likely structure when it occurs amid other error signals in the sentence. In essence, the input children hear is full of unexpected words, so the word that rules out an incorrect initial parse is drowned out.

It is worth noting that the eye-tracking studies that reveal children's difficulty with garden-path sentences use only a small set of objects. Children see the objects that will be referenced in the sentences they hear, either as toys on small platforms in front of them, or as computer images. The very act of displaying these images for children may reduce children's surprise at hearing them, even for relatively infrequent nouns. However, the visual-world eye-tracking paradigm is robustly sensitive to lexical frequency effects, even for young children (Magnuson et al., 2003; Dahan et al., 2007; Borovsky et al., 2016). Participants are faster to look to higher frequency words than they are to look to low-frequency ones. Seeing visual representations of the upcoming nouns therefore does not make children expect all of them to be equally plausible, and it is reasonable to assume that children's surprise at hearing unexpected nouns will still be reflected in the eye-tracking studies that demonstrate Kindergarten-path effects.

At a minimum, the present findings indicate that statistical properties of children's input could play a role in their relatively impaired performance on garden-path sentences. This is not to say that input effects are mutually exclusive with a role for the development of cognitive-control; given the preponderance of evidence that children's executive function skills are still developing (Davidson et al., 2006; Diamond et al., 2002; Müller et al., 2005), it is premature to try to rule out the cognitive-control account. Input differences may interact with children's still-developing cognitive-control abilities: children may have additional difficulty revising their initial incorrect parsing commitments because the error signal they receive is weaker. Even if children and adults were equally adept at conflict *resolution*, these results suggest that children are at a disadvantage for detecting the conflict signal to begin with. Indeed, children may be disadvantaged in two different ways: the speech they hear doesn't prepare them to detect this type of syntactic ambiguity, and when they do detect it, they have trouble giving up their initial parse in favor of the correct one.

In order to fully spell out this account, further work is needed to make sure the current results remain consistent. Since the Kindergarten-path effect holds across languages (Choi & Trueswell, 2010; Huang et al., 2013; Omaki et al., 2014; Lassotta et al., 2016), similar results are expected for non-English corpora of child-directed speech. Additionally, by the time children are eight years old, the Kindergarten-path effect with sentences like (1) nearly disappears (Weighall, 2008), and so we might expect the properties of speech to these slightly older children to reflect this change. Additionally, PP-attachment ambiguity is only one type of Kindergarten-path sentence. Children have been found to have non-adult-like interpretations for a variety of other temporarily ambiguous structures such as passives and relative-clause attachment ambiguity (Huang & Arnold, 2016; Traxler, 2007; Huang et al., 2013, 2017). Future work will test a wide variety of temporarily ambiguous sentences to see if this finding can explain children's performance on other structures. If low surprisal values are consistently found where children fail to notice disambiguation in other sentence types, this would suggest that differences in the way we speak to children and adults can account for a substantial portion of children's difficulty in navigating syntactic ambiguity.

Acknowledgments

We would like to thank audiences at the University of Maryland's Language Science Lunch Talk series, the UMD Probabilistic Modeling Reading Group and the UMD Language & Cognition Lab for their helpful feedback. This work was supported by National Science Foundation Grants #DGE-1449815 and #DGE-1322106.

References

Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, *137*(2), 181-189.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child development*, *87*(6), 1893-1908.

Choi, Y., & Trueswell, J. C. (2010). Children's (in) ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*(1), 41-61.

Dahan, D., Tanenhaus, M. K., & Salverda, A. P. (2007). The influence of visual processing on phonetically driven saccades in the "visual world" paradigm. In *Eye Movements* (pp. 471-486). Elsevier.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*(3), 352.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the ACM*, *13*(2), 94-102.

Eriksen, B. A., & Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143-149.

Ferraro, F., & Van Durme, B. (2016). A Unified Bayesian Model of Scripts, Frames and Language. In AAAI (pp. 2601-2607).

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS One*, *8*(10), e77661.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178-210.

Futrell, R., Levy, R., & Gibson, E. (2017). Generalizing dependency distance: Comment on "Dependency distance: A new perspective on syntactic patterns in natural languages" by Haitao Liu et al. *Physics of Life Reviews*, *21*, 197-199.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 312–7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129-153.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992, March). SWITCHBOARD: Telephone speech corpus for research and development. In *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 517-520). IEEE.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*. Association for Computational Linguistics.

Huang, Y. T., & Arnold, A. R. (2016). Word learning in linguistic context: Processing and memory effects. *Cognition*, *156*, 71-87.

Huang, Y. T., Leech, K., & Rowe, M. L. (2017). Exploring socioeconomic differences in syntactic development through the lens of real-time processing. *Cognition*, *159*, 61-75.

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, *69*(4), 589-606.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*(6), 597-626.
Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, *20*(3), 573-591.
Laakso, A., & Smith, L. B. (2007). Pronouns and verbs in adult speech to children: A corpus analysis. *Journal of Child Language*, *34*(4), 725-763.
Lassotta, R., Omaki, A., & Franck, J. (2016). Developmental changes in misinterpretation of garden-path wh-questions in French. *The Quarterly Journal of Experimental Psychology*, *69*(5), 829-854.
Lieven, E. (1984). Interactional style & children's language learning. *Topics in Language Disorders*, *4*(4), 15-23.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.
Luong, M. T., Frank, M. C., & Johnson, M. (2013). Parsing entire discourses as very long strings. *Transactions of the Association for Computational Linguistics*, *1*, 315-326.
MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk.* Mahwah, NJ: Lawrence Erlbaum Associates.
Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202.
Mazuka, R., Jincho, N., & Oishi, H. (2009). Development of executive control and language processing. *Language and Linguistics Compass*, *3*(1), 59-89.
Müller, U., Zelazo, P. D., & Imrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation?. *Cognitive Development*, *20*(2), 173-189.
Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Ferguson & C. Snow (Eds.), *Talking to children*. New York: Cambridge Univ. Press.
Novick, J., Trueswell, J. C., & Thompson-Schill, S. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5*(3), 263-281.
Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development*, *10*(3), 206-233.
Pearl, L., & Sprouse, J. (2013). Computational models of acquisition for islands. *Experimental syntax and islands effects*, 109-131.
Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths. *Journal of Experimental Psych.: Learning, Memory, and Cognition*, *24*(4), 940.
Qi, Z., Love, J., Fisher, C., & Brown-Schmidt, S. (2020). Referential context and executive functioning influence children's resolution of syntactic ambiguity. *Journal of Experimental Psych.: Learning, Memory, and Cognition.* Retrieved from https://psyarxiv.com/jc92m/
Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358-374.
Richards, B. J. (1994). Child-directed speech and influences on language acquisition: Methodology and interpretation.
Snow, C. E., and Ferguson, C. A. (eds.) (1977) *Talking to children: language input and acquisition.* Cambridge: Cambridge Univ. Press.
Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, *21*(2), 165-201.
Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory & Language*, *40*, 136-150.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.
Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs. *Journal of Child Language*, *24*(3), 535-565.
Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, *35*(5), 1107-1121.
Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89-134.
Warren-Leubecker, A., & Bohannon III, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development*, *55*(4), 1379-1385.
Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*(2), 75-95.
Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, *141*, 187-209.