## Title

Multiword Units Predict Non-inversion Errors in Children's Wh-questions: "What Corpus Data Can Tell Us?"



**Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

**Authors**

McCauley, Stewart M.
Bannard, Colin
Theakston, Anna
et al.

**Publication Date**

2019

Peer reviewed

# Multiword Units Predict Non-inversion Errors in Children's *Wh*-questions: "What Corpus Data Can Tell Us?"

**Stewart M. McCauley (stewart-mccauley@uiowa.edu)**
Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242

**Colin Bannard (Colin.Bannard@liverpool.ac.uk)**
Department of Psychological Sciences, University of Liverpool, Liverpool, UK, L3 5TR

**Anna Theakston (anna.theakston@manchester.ac.uk)**
**Michelle Davis (michelle.davis@manchester.ac.uk)**
Division of Human Communication, Development & Hearing, University of Manchester, Manchester, UK, M13 9PL

**Thea Cameron-Faulkner (T.Cameron@manchester.ac.uk)**
Department of Linguistics and English Language, University of Manchester, Manchester, UK, M13 9PL

**Ben Ambridge (Ben.Ambridge@liverpool.ac.uk)**
Department of Psychological Sciences, University of Liverpool, Liverpool, UK, L3 5TR

**Abstract**

Subject-auxiliary inversion in interrogatives has been a topic of great interest in language acquisition research, and has often been held up as evidence for the structure-dependence of grammar. Usage-based and nativist approaches posit different representations and processes underlying children's question formation and therefore predict different causes for these errors. Here, we explore the question of whether input statistics predict children's spontaneous non-inversion errors with *wh-* questions. In contrast to previous studies, we look at properties of the non-inverted, errorful forms of questions. Through a series of corpus analyses, we show that the frequency of uninverted subsequences (e.g., "*she is going"* in "*what she is going to do?*"*) is a good predictor of children's errors, consistent with recent evidence for multiword units in children's comprehension and production. This finding has implications for the types of mental representations and cognitive processes researchers ascribe to children acquiring a first language.

**Keywords:** language acquisition; interrogatives; corpora; corpus analyses; usage-based approach; chunking

## Introduction

Whether the input available to children is sufficient to explain their emerging language abilities is a fundamental question in cognitive science (Chomsky, 1957; Skinner, 1957). Central to the ongoing discussion are tensions between the view of grammar as the result of gradual abstraction over the input (e.g., Lieven, Salomo & Tomasello, 2009; Tomasello, 2003), and approaches in which the acquisition process is guided by innate, language-specific biases (e.g., Pinker, 1999; Fisher, 2002).

In the realm of theoretical linguistics, work in support of the latter approach has focused on specific linguistic phenomena, such as interrogatives. A topic of particular interest is that of subject-auxiliary inversion, which has been held up as evidence for the structure-dependence of grammar (e.g., Crain, 1991; Berwick, Pietroski, Yankama, & Chomsky, 2011), and is often still discussed in the same terms as it was half a century ago (Chomsky, 1968).

In developmental psycholinguistics, a great deal of work has also focused on interrogatives, in part because they represent some of the few sentence types for which English-speaking children reliably make errors involving word order (e.g., Klima & Bellugi, 1966; Stromswold, 1990). Moreover, these sentence types provide a means to evaluate subject-auxiliary inversion as evidence for structure-dependence within a developmental framework. This applies to *wh*-questions especially: as both the *wh*-word and the auxiliary are fronted, it has been argued that they are structurally more complex than *yes/no* questions (e.g., Pozzan & Valian, 2017; Jakubowicz, 2011); and unlike *yes/no* questions, children rarely encounter *wh*-questions in uninverted form as part of the input, yet still make errors of uninversion as in (1).

(1) *What <u>they are</u> doing over there ? **

Thus, *wh-* questions represent an ideal case for mediating between nativist and constructionist approaches, as each posit different representations and processes underlying children's errors and therefore predict different error properties. While the former emphasizes abstract structural considerations, the latter perspective stresses the importance of input frequency in supporting lexically-specific representations.

In line with structure-dependence accounts, a number of researchers have argued for earlier acquisition of argument *wh*-questions than adjunct *wh*-questions, based on their structural properties (e.g., Stromswold 1990, de Villiers 1991). Consistent with this, Pozzan and Valian (2017) report higher non-inversion rates for adjunct than for argument *wh-* questions, a finding they argue to be independent of input frequencies (as might be predicted under usage-based approaches). However, frequency is not rigorously controlled for in the design of the stimulus items themselves, nor is the frequency of substrings beyond the *wh*-word/auxiliary combination considered (in the following subsection, we discuss why this may be of importance).

Initial support for usage-based approaches to subject-auxiliary inversion came from a corpus analysis of one child's early *wh*- questions (Rowland & Pine, 2000). The authors found that the frequency of specific *wh*-word + auxiliary combinations reliably predicted non-inversion rates. Ambridge, Rowland, Theakston, and Tomasello (2006) extended this finding with an elicited production study in which *wh*-word + auxiliary combinations predicted non-inversion rates in children aged 3;6 to 4;6. Moreover, *wh*-word alone was not found to predict errors, in contrast to structure-dependence accounts (e.g., Pozzan & Valian, 2017). Rather, the pattern of results was consistent with the notion of lexically-specific representations driving performance with particular question types.

In a further elicitation study, Ambridge and Rowland (2009) investigated a wider range of question types, including negative polarity questions, replicating the finding that *wh*-word + auxiliary frames predicted error rates. Though the relevant frequency dimensions were not controlled for in a rigorous way, Ambridge and Rowland also found initial support for the notion that patterns learned from declarative utterances may also shape errors. It is to this possibility that we turn in the present study.

## A Role for Multiword Units in Predicting Non-inversion Errors

A serious limitation of previous work on subject-auxiliary inversion is that only the distributional properties of *correct* forms have been taken into account. This partly stems from the lingering influence of theoretical frameworks in which individual words are viewed as the fundamental units over which language processing take place (e.g., Pinker, 1999). After all, the correctly inverted and errorful, non-inverted forms of a question contain the same set of words; only the word order differs. Thus, if words are the fundamental units of language, we would not expect the distributional properties of an errorful form to play a role in question formation.

Recent years, however, have seen an explosion of psycholinguistic data suggesting that language users are not only sensitive to the properties of compositional multiword sequences, but—in some sense—store and actively utilize such sequences in comprehension and production, as linguistic units in their own right. The frequency of such multiword units—or "chunks"—has been shown to facilitate processing in adult comprehension (e.g., Arnon & Snider, 2010; Bannard, 2006; Reali & Christiansen, 2007) as well as production (e.g., Janssen & Barber, 2012). These findings have received further support from event-related brain potentials (Tremblay & Baayen, 2010) and eye-tracking data (Siyanova-Chanturia, Conklin, & van Hueven, 2011).

Importantly, these findings are mirrored in psycholinguistic work with children (see Theakston & Lieven, 2017 for an overview). Bannard and Matthews (2008) found that, when controlling for substring frequency, overall sequence frequency predicted the speed and accuracy with which 2- and 3-year-olds produced compositional phrases. Arnon and Clark (2011) report evidence that multiword chunk frequency intersects with morphological development: errors of noun plural overregularization were significantly reduced when irregular plurals were produced in the context of more frequent sequences. Moreover, multiword units exhibit the same type of age-of-acquisition (AoA) effects as do individual words, when AoA is determined by either subjective ratings or by corpus-based metrics (Arnon, McCauley, & Christiansen, 2017). Taken together, these findings underscore the possibility that multiword chunks serve as building blocks for language learning.

The importance of these findings to more general theoretical debates is further highlighted by computational modeling work which has shown that abstraction over stored sequences can lead to a considerable amount of linguistic productivity (e.g., Solan, Horn, Ruppin, & Edelman, 2005). Even models lacking abstraction have served to demonstrate that associative learning of chunks from naturalistic input can account for a substantial portion of children's language production (McCauley & Christiansen, 2019).

Therefore, if children are sensitive to the properties of multiword sequences, we might expect such information to play a role in *wh*-question formation. Take, for instance, the following correctly inverted and non-inverted (errorful) forms (2-3):

(2) *What <u>is she going</u> to do ?*

(3) *What <u>she is going</u> to do ? **

If the uninverted strings "*she is going*" or "*is going*" are highly frequent in the child's input, we might expect—given evidence that multiword chunks play a role in learning and processing—that the child is more likely to produce the errorful form. By the same token, we might expect the frequency of "*is she going*" or "*she going*" to alter this likelihood in the opposite direction. From this perspective, chunks from both the correctly inverted and non-inverted forms might be seen as competing. In other words, multiword sequence frequencies from the correctly inverted and non-inverted forms are both important, insofar as they relate to one another.

## The Present Study

If such a relationship exists at all, it is likely to be a complex one, mediated by a host of distributional, pragmatic, and semantic factors. In the present study, we take an initial step towards disentangling these factors by considering, simultaneously, the many distributional factors at play. Not only have the frequencies of individual *wh*-words and auxiliaries been argued to shape errors, but also the frequencies of distinct *wh*-word/auxiliary combinations themselves (e.g., Rowland & Pine, 2000). Given the perspective we have put forth regarding a role for multiword sequences stretching beyond the *wh*-word and auxiliary, it is necessary to consider the distributional properties of individual words and higher-order *n*-grams for both the correctly inverted and uninverted forms of questions, simultaneously.

In the present study, we evaluate the role of multiword units in early *wh*-question production by using distributional statistics from child-directed speech to predict children's *spontaneous* uninversion errors. Using the entire English portion of the CHILDES database (MacWhinney, 2000), we collect distributional statistics for words and higher-order *n*-grams, which are then used to construct a logistic regression model of children's correctly inverted and errorful (uninverted) questions across the 12 most question-rich corpora. Thus, we are able to test whether, and to what extent, frequencies for individual words and multiword combinations predict spontaneous error rates. Moreover, this allows us to evaluate the role played by multiword sequences from the uninverted forms of questions while controlling for the statistics of the correctly inverted forms, and vice-versa.

In this context, usage-based approaches make predictions that are separable and distinct from those made by theories emphasizing abstract, system-wide principles: if children are forming questions based on structural properties, we would not expect to see a role for uninverted *n*-gram statistics in predicting uninversion errors. Moreover, we would expect structural differences in question type (e.g., argument questions vs. adjunct questions) to be better predictors of correct inversion than frequency (e.g., Pozzan & Valian, 2017). By contrast, usage-based approaches would predict experience with particular *wh*-words, auxiliaries, and even specific subjects/verbs to be robust predictors of error rates, and would quite naturally accommodate findings that *n*-gram sequences from the uninverted forms predict error rates. Under such a view, abstract grammatical constructions tied to questions would emerge gradually as a process of abstracting over stored sequences, and this would be reflected in the probabilities with which children fail to correctly invert certain sentences.

## Methods

The corpus analysis consisted of three general phases: extraction of all child-produced *wh-* questions from a set of target corpora, followed by semi-automated identification of uninversion errors; collection of *n*-gram statistics for child-directed speech in English; and mixed-effects logistic regression modeling to determine which *n*-gram statistics predicted uninversion errors in the extracted questions.

### Corpus Selection and Preparation

We began by extracting the 12 corpora with the highest number of *wh-* questions from the English language portion of the CHILDES database (MacWhinney, 2000). Each corpus followed a single target child and spanned at least one year of development; the age range and nationality for each target child is shown in Table 1 alongside citation information.

**Table 1:** Details of CHILDES Corpora Used in Analysis of Uninversion Errors

| Target Child | Corpus | Age Range |
|---|---|---|
| Abe | Kuczaj, 1977 | 2;04-5;00 |
| Adam | Brown, 1973 | 2;03-5;02 |
| Eleanor | Lieven et al., 2009 | 2;00-3;00 |
| Ethan | Demuth & McCullough, 2009 | 0;11-2;11 |
| Fraser | Lieven et al., 2009 | 2;00-3;01 |
| Laura | Braunwald, 1976 | 1;05-7;00 |
| Lara | Rowland & Fletcher, 2006 | 1;09-3;03 |
| Lily | Demuth & McCullough, 2009 | 1;01-4;00 |
| Naima | Demuth & McCullough, 2009 | 0;11-3;10 |
| Ross | MacWhinney, 1991 | 1;04-7;08 |
| Sarah | Brown, 1973 | 2;03-5;01 |
| Thomas | Maslen et al., 2004 | 2;00-4;11 |

Each corpus was then prepared for analysis using an automated procedure which removed codes, tags, and punctuation, leaving only speaker identifiers and the original sequence of words. Lines consisting solely of morphological tags (included as standard in CHILDES corpora) were unaffected by this procedure and were retained for later use in extracting uninversion errors.

As part of this procedure, contractions were split into their component words: e.g., "what's he doing" was re-coded as "what is he doing." As corpus annotation differs in terms of how contractions are transcribed (leading to arbitrary noise), this step ensured that modeling work reflected accurate *n*-gram frequencies for *wh-* words and auxiliaries across all questions. As a further step we collapsed the pronouns "she" and "he" into a single form to control for individual differences across children's exposure to gender pronouns.

### *Wh-* Question and Uninversion Error Candidate Extraction and Coding

Child-produced *wh-* questions were automatically extracted from the target corpora by utilizing the standard default morphological tagging included in CHILDES. All extracted questions featured a *wh-* word in the first position, followed immediately by an auxiliary. This yielded approximately 13,000 child-produced *wh-* questions across the 12 corpora.

For the purpose of automatically identifying possible uninversion errors, we extracted, from the full corpora, all child questions which featured a *wh-* word in the initial position which was not immediately followed by an auxiliary. These candidate items were then manually coded for error type by the first author, yielding a total of 300 identified uninversion errors produced across the target children. *wh-* questions featuring an error type other than uninversion (such as doubling or omission errors) were excluded from our dataset. Importantly, our analyses were restricted to questions produced before the age of five years.

## *N*-gram Data Collection

In order to capture *n*-gram statistics which accurately reflected the nature of child-directed speech in the English language, we gathered *n*-gram frequencies for the entire English (UK and US) portion of the CHILDES database. This allowed us to overcome issues of data sparseness arising from corpus size (Manning & Schütze, 1999).

The aggregated corpus was prepared for data collection following the same procedure described in the above subsection. Frequencies were then collected for unigrams (single words), bigrams (word pairs), and trigrams (word triplets), which were then applied to each of the *wh*-questions extracted for the 12 target child corpora. To this end, *n*-gram statistics were calculated for each question (separate unigram counts for each word, separate bigram counts for each word pair, and so forth). Thus, for the question "what is that," three unigram counts (one for each of three word positions), two bigram counts (one for each of two word pair positions), and one trigram count (for the single word triplet position) were available.

Because our statistical analyses aimed to explore the role of multiword chunk frequency in shaping children's uninversion errors, we sought to directly compare the correctly inverted "target question" for children's uninversion errors to the correctly inverted questions which made up the rest of the dataset. To achieve this, we calculated *n*-gram frequencies for the correctly inverted forms of the uninverted questions identified by the earlier procedure. Uninversion errors were "corrected" by hand in order to achieve this.

By the same token, we also sought to explore the role of multiword sequence frequencies for the relevant uninverted question forms in determining error rates. For this, we retained the original child uninversion errors and employed an automated procedure to produce the errorful, uninverted form corresponding to each correctly inverted question in the corpus. The second and third words could not simply be swapped because a large number of questions featured multiword subject noun phrases, such as "where is my red ball?" Thus, to automatically achieve a realistic uninverted form across such a large number of questions, we first chunked utterances using a shallow parser (Punyakanok & Roth, 2001). Shallow parsers are widely used tools in the field of natural language processing which segment out the non-overlapping, non-embedded phrases in a text. For instance, the shallow parser output for the previous example would be: "[where] [is] [my red ball]." After submitting all correctly inverted questions to the shallow parser, we merely switched the second and third chunks, yielding the relevant, uninverted errorful forms, such as "where my red ball is?"

Thus, we collected unigram, bigram, and trigram statistics for each position across all correctly inverted questions (and, in the case of uninversion errors, the correctly inverted target questions), alongside a separate set of *n*-gram statistics for the uninversion errors (and, in the case of correctly inverted questions, the relevant errorful form).

## Analysis

In order to evaluate the predictive relationship between multiword chunk frequency and uninversion errors, we used mixed-effects logistic regression modeling (cf. Agresti, 2002). We carried out a set of model comparisons to determine which *n*-gram frequencies were uniquely predictive of the relationship. This involved selecting predictors at each *n*-gram level separately, starting at the unigram level before moving to the bigram level, followed by the trigram level.

Questions originally produced by the target children in their correctly inverted form were coded as 0, while questions produced in an errorful, uninverted form were coded as 1. *N*-gram frequencies were then used as predictors for this binary variable. All models included a random intercept for child, to reflect the fact that the 12 target children may differ in the extent to which their errors could be predicted by *n*-gram frequencies. By-child random slopes were also included where they improved fit.

Our model comparisons sought to evaluate *n*-gram frequencies of both the correctly inverted question and their corresponding uninverted (errorful) forms as predictors of child uninversion error. The model comparison procedure was designed such that the risk of false positives for higher-order *n*-grams was insignificant, as we conservatively prioritized lower-order *n*-grams in the selection process. Importantly, all predictors were log-transformed and scaled. All model comparisons were carried out using log-likelihood ratio tests.

Starting at the unigram level, we used a leave-one-out procedure to determine which predictors explained variance over and above that explained by any other variable. The full baseline model at this level included random effects of the first 5 unigrams (by child) as well as fixed effects for these 5 unigrams. This was then compared to five subsequent models, each leaving out the fixed effect term for a different unigram (random effects by child were included for every unigram in each model). Removal of only the first two unigrams harmed model fit to a significant extent, according to log-likelihood tests. Thus, these two unigrams were held over for the next level of model comparisons.

The same procedure described for unigrams was then carried out for the first four bigrams, but with random (by child) and fixed effects for the first two unigrams also included in each model (as unigrams are identical across the inverted and uninverted forms, only one set was included in the previous step). Importantly, bigrams from both the correctly inverted and the corresponding errorful forms were included at this second step.

For correctly inverted question forms, removal of the third and fourth bigrams harmed model fit to a statistically significant extent, according to the log-likelihood tests, while for the uninverted forms, removal of the second, third, and fourth bigrams harmed model fit. Thus, in addition to the first two unigrams from the previous step, the third and fourth bigrams from the correctly-inverted question forms and the second, third, and fourth bigrams from the errorful (uninverted) forms were held over for the final set of model comparisons.

For the first three trigrams, the same procedure was followed once more (with random and fixed effects for the first two unigrams and first two bigrams). Only removal of the second and third trigrams from the uninverted/errorful question forms harmed model fit to a significant extent.

Thus, the final set of predictors included the first two unigrams, the third and fourth bigrams from the correctly inverted forms, the second, third, and fourth bigrams from the uninverted forms, and the second and third trigrams from the uninverted forms.

## Results

Our model comparison procedure (as described above) yielded a model with 9 *n*-gram predictors: the first two unigrams, third and fourth bigrams from the correctly inverted question forms; and the second, third, and fourth bigrams as well as the second and third trigrams for the errorful (uninverted) question forms. The log-likelihood, chi-squared value, and p-value for each model comparison is shown in Table 2.

**Table 2:** Results of Model Comparisons

| **Left-out Predictor** | **Log-likelihood** | $\chi^2$ | ***p*-value** |
|---|---|---|---|
| Unigram (full/baseline) | -702.13 | - | - |
| Unigram 1 | -705.6 | 6.95 | 0.00 ** |
| Unigram 2 | -707.16 | 10.07 | 0.00 ** |
| Unigram 3 | -702.27 | 0.29 | 0.59 |
| Unigram 4 | -702.13 | 0.00 | 0.97 |
| Unigram 5 | -702.20 | 0.14 | 0.71 |
| Bigram (full/baseline) | -626.40 | - | - |
| Bigram 1 | -627.28 | 1.76 | 0.19 |
| Bigram 2 | -627.20 | 1.59 | 0.21 |
| Bigram 3 | -631.41 | 10.01 | 0.00 ** |
| Bigram 4 | -632.68 | 12.55 | 0.00 *** |
| Trigram (full/baseline) | -614.62 | - | - |
| Trigram 1 | -615.44 | 1.641 | 0.2002 |
| Trigram 2 | -615.69 | 2.141 | 0.1434 |
| Trigram 3 | -614.67 | 0.103 | 0.748 |
| Uninverted Bigram (full/baseline) | -626.40 | - | - |
| Uninverted Bigram 1 | -626.42 | 0.02 | 0.88 |
| Uninverted Bigram 2 | -634.79 | 16.77 | 0.00 *** |
| Uninverted Bigram 3 | -634.87 | 16.94 | 0.00 *** |
| Uninverted Bigram 4 | -632.5 | 12.19 | 0.00 *** |
| Uninverted Trigram (full/baseline) | -614.62 | - | - |
| Uninverted Trigram 1 | -614.87 | 0.505 | 0.4772 |
| Uninverted Trigram 2 | -617.55 | 5.874 | 0.02 * |
| Uninverted Trigram 3 | -618.41 | 7.582 | 0.01 ** |

To help understand the relationship of these *n*-gram frequencies with child uninversion errors, we constructed non-partial (single-predictor) models for each of the final variables, as reported in Table 3. Each model included a random intercept for target child and a random effect (by child) for the relevant predictor as well as the fixed effect. This procedure was preferred as, in a multi-predictor model, estimates may change sign based on the relative strength of predictor correlations with the dependent variable (cf. Wurm & Fisicaro, 2014).

The first and second unigram frequencies (corresponding to the *wh-* word and auxiliary) were significant predictors with negative estimates, indicating lower likelihood of an uninversion error with more frequent items. Importantly, for higher-order *n*-gram predictors drawn from the errorful, uninverted question forms, the estimate was positive. This means that the higher the *n*-gram frequency was for the uninverted form of a question, the more likely it was for that question to have been produced in its uninverted form.

**Table 3:** Results of Non-partial Models

| ***N*-gram** | **β** | **Std. Error** | ***Z*** | ***p*-value** |
|---|---|---|---|---|
| Uni 1 | -0.792 | 0.27 | -2.91 | 0.004 ** |
| Uni 2 | -0.634 | 0.11 | -5.34 | 0.000 *** |
| Bi 3 | 0.031 | 0.11 | 0.25 | 0.795 |
| Bi 4 | 0.239 | 0.14 | 1.64 | 0.100 |
| Bi 2 (*uninv.*) | 0.328 | 0.11 | 2.89 | 0.004 |
| Bi 3 (*uninv.*) | 0.563 | 0.13 | 4.24 | 0.000 *** |
| Bi 4 (*uninv.*) | 0.207 | 0.16 | 1.26 | 0.207 |
| Tri 2 (*uninv.*) | 0.462 | 0.10 | 4.44 | 0.000 *** |
| Tri 3 (*uninv.*) | 0.454 | 0.11 | 4.03 | 0.000 *** |

## General Discussion

The corpus analyses presented here represent, to our knowledge, the most rigorous attempt to control for input frequency in analyzing non-inversion errors to date. We find that, when *n*-gram frequencies from both the correctly-inverted, “target” form of a question, and the non-inverted, “errorful” form of a question are considered in parallel, frequency is a robust predictor of when non-inversion errors will occur. Moreover, the frequencies of higher-order *n*-

grams from the non-inverted form are shown to be more robust predictors than frequencies from the correctly inverted form.

This finding appears to stem from children's use of multiword units in production (e.g., Bannard & Matthews, 2008). Consider the effect of the (non-inverted) second trigram in the context of the following non-inversion error: "*where* <u>*we can go*</u> *today?**" The more heavily *we can go* holds together as a unit in the child's language experience, the less likely the child will be to break up the sequence by fronting the auxiliary *can* (e.g., by relying on a lexical frame for *what can*). Similar reasoning can be applied to the effect of the non-inverted third bigram (*can go,* in this example). Errors caused by the intrusion of overlearned sequences occur in all kinds of human action (Bannard et al., in press).

Thus, our findings weigh in favor of previous proposals that children rely on lexically-based representations in question formation (e.g., Rowland & Pine, 2000) and support the proposal that material learned from declarative utterances can drive systematic errors (Ambridge & Rowland, 2009). Our findings are inconsistent, however, with structure-dependent accounts of children's *wh*-questions (e.g., de Villiers, 1991).

The present study, therefore, offers an interesting additional line of evidence supporting usage-based approaches, especially accounts of language development which stress the importance of multiword units (e.g., Theakston & Lieven, 2017; McCauley & Christiansen, 2019) including exemplar-based approaches (Ambridge, 2018).

## References


Ambridge, B. (2018, July 25). Against stored abstractions: A radical exemplar model of language acquisition. https://doi.org/10.31234/osf.io/gy3ah

Ambridge, B., Rowland, C.F., Theakston, A.L., & Tomasello, M. (2006). Comparing different accounts of inversion errors in children's non-subject *wh*-questions: "What experimental data can tell us?." *Journal of Child Language, 33*, 519-557.

Ambridge, B., & Rowland, C.F. (2009). Predicting children's errors with negative questions: Testing a schema-combination account. *Cognitive Linguistics, 20*, 225-266.

Agresti, A. (2002). *Categorical data analysis.* Hoboken, NJ: Wiley.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multiword phrases. *Journal of Memory and Language, 62*, 67–82.

Arnon, I., McCauley, S.M. & Christiansen, M.H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language, 92*, 265-280.

Bannard, C. (2006). *Acquiring phrasal lexicons from corpora*. Doctoral dissertation: University of Edinburgh.

Bannard, C., Leriche, M., Bandmann, O., Brown, C., Ferracane, E., Sánchez-Ferro, A., Obeso, J., Redgrave, P., & Stafford, T. (in press). Reduced habit-driven errors in Parkinson's Disease. *Nature Scientific Reports.*

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning. *Psychological Science, 19*, 241.

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science, 35,* 1207-1242.

Chomsky, N. (1957). *Syntactic structure.* Hague: Mouton.

Chomsky, N. (1968). *Language and mind.* New York: Harcourt, Brace, Jovanovitch.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14,* 597-612.

de Villiers, J. (1991). Why question? In T. L. Maxfield & B. Plunkett (eds.), *Papers in the acquisition of wh: Proceedings of the UMASS Roundtable*, May 1990. Amherst, MA: University of Massachusetts Occasional Papers.

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition, 82*, 259-278.

Jakubowicz, C. (2011). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua, 121*, 339-351.

Klima, E.S., & Bellugi, U. (1966). Syntactic regularities in children's speech. In J. Lyons & R. Wales (Eds.), *Psycholinguistic papers* (pp. 183-208). Edinburgh: Edinburgh University Press.

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20,* 481-507.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge: MIT press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, Volume II: The Database.* Mahwah, NJ: Lawrence Erlbaum Associates.

McCauley, S.M. & Christiansen, M.H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review, 126,* 1-51.

Pinker, S. (1999). *Words and rules: The ingredients of language.* New York: Basic Books.

Pozzan, L. & Valian, V. (2017). Asking questions in child English: Evidence for early abstract representations. *Language Acquisition, 24,* 209-233.

Punyakanok, V., & Roth, D. (2001). The use of classifiers in sequential inference. In *Proceedings of NIPS 2001* (pp. 995-1001).

Reali, F. & Christiansen, M.H. (2007). Word-chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology, 60*, 161-170.

Rowland, C.F. & Pine, J.M. (2000). Subject–auxiliary inversion errors and *wh*-question acquisition: 'What children do know?*' Journal of Child Language, 27*, 157–81.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*, 251-272.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences, 102*, 11629-11634.

Stromswold, K. J. (1990*). Learnability and the acquisition of auxiliaries*. Doctoral dissertation: Massachusetts Institute of Technology.

Tomasello, M. (2009). *Constructing a language.* Cambridge: Harvard University Press.

Theakston, A. & Lieven, E. (2017), Multiunit sequences in first language acquisition. *Topics in Cognitive Science, 9*, 588-603.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences. In D. Wood (Ed.) *Perspectives on formulaic language* (pp. 151-173). London: Continuum International Publishing Group.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language, 72*, 37-48.