Proposal Outline

Jul 10, 2025

Thomas Morton

0 Just drop the subject: A controlled rearing study of Subject Drop in English

It's the privilege of linguistics to look at complex phenomena in language and ask how children acquire such behavior. Along the way to adult performance, children exhibit linguistic behavior unlike the language they are learning. Before mastering some phenomena, children's language can exhibit behavior similar to other languages in content. The question becomes, do children exhibit linguistic behavior unlike their target language because they have yet to settle on the single target generalization, or that they have learned the target language but processing constraints shape linguistic outputs in ways that appear outside of distribution.

A classic case of this appears in the study of children's acquisition of non-null-subject languages. In these languages[1], like English, omitting a subject pronoun is unacceptable in nearly all cases, unlike Italian which allows for dropped subjects in many, if not most, cases. In Languages like Italian, children tend to preference towards omitting subjects rather quickly, the optional preference seeming fairly easy to learn. On the other hand, English children persist in omitting subjects (despite being illicit in their language) up until nearly three years old. Although English-speaking children drop their subjects at rates much lower than Italian-speaking children, the appearance of subject dropping in English children begs the question whether they content-fully believe they speak a language that allows for subject-dropping; or, despite knowing such a rule that dropped subjects are illicit in English, such artifacts appear as a result of their language processor. In this case a processor that omits subjects as a result of capacity constraints or economic principles overtaking grammatical principles.

There is a rich literature investigating both sides of this question, one side investigating the path that children taken in acquiring grammatical generalizations and another that looks to explain the artifacts of early learning as processing effects. The study proposed here looks to ask these questions again, using Large

[1] Although this is probably not a high-fidelity splitting of language, it's a rough go at grouping them

Language Models (LLMs) as a new tool in the psycholinguistic arsenal for asking questions about the learning and processing of language. Specifically, we wish to investigate what role specific sources of linguistic evidence contribute to the direct and indirect learning of the English overt subject constraint.

LLMs are specifically well-served to ask these questions, as it is prohibitively difficult to know what a child's linguistic input looks like, or even more to manipulate the kind of input available to children while investigating their learning. In this way, we can manipulate the sources of information available to models and compare the causal effect that different sources of linguistic information (or their lack there-of) have on acquire linguistic generalizations.

Further, in asking questions of children's linguistic processing, exposing children to linguistic stimuli longitudinally to test changes in children's performance would surely introduce confounds of exposure and learning. Meanwhile, LLMs offer us the ability to sample a model's performance throughout training without influencing future performance of the model. Further, we can investigate the model on a wider range of evaluation stimuli without worrying about effects of fatigue.

The goal then, is to investigate large language models as candidate learners of the overt subject constraint and compare theories of learning and processing to see which theories best capture the performance of models when learning and processing sentences with and without subjects.

0.0 Accounts of subject drop

Subject drop in english children's early language has been widely attested in early children's corpus work [1, 2]. Examples like:

(1) Shake hands.
Turn light off.
Want go get it.
Show mommy that.
Now making muffins

Bloom [1] put forward the claim that English children drop their subjects for performance reason. Bloom noted that for the verb *make* when children produce a subject, they are also more likely to omit the verb. Further, more subjects were dropped in sentences with negation. Under this account, when children encounter contexts with heavy cognitive load, they are prone to omit information that may be less important or otherwise contextually recoverable. Evidence for such cases come from cases where children from even very early ages, before when they could have possibly acquired such a rule, show distributional features of their target language. For example Valian [3] reports that English children still produce substantially more overt subjects than Italian children before averaging over a mean length utterance of 2. One caveat to this account is the asymmetry between subject and object pronouns, such that subject pronouns are much more likely to be dropped than object pronouns.

[4] proposes that this can be contended for if such processing asymmetries can be found in orthogonal contexts. For example, longer names are omitted more than shorter names, across both the subject and object positions. The case isn't necessarily that such accounts are not about learning, except that there is some learning process that is occurring during children's development that interacts with this processing element and that accounts of this effect that are purely grammatical in explanation lose out explanatorily. The idea that our linguistic output is determined by the interplay between resource-limited processors guided by more abstract representations is not a new one [5].

They propose that there must be some particular processing difficulty in speaking a subject as compared to speaking the object. This kind of account is criticized by Hyams and Wexler [6, 7] argue that there is not sufficient evidence to suggest that there should be a difference in the processing of a subject that should lead to such start assymetries between the subject and object position. They suggest that children are not considering grammars where objects can be dropped, but they are where subjects can be dropped, and so this is an artifact of their competence, not performance.

For example, one assumption that must be made by a performance account allowing for this asymmetry is subjects are inherently more difficult to produce than objects, whereas work in the area of sentence production would suggest that in-fact the subject should require less work to process, because it can be planned separately from the verb [8]. Further, initiating speech altogether requires fairly little planning, with some experimental evidence suggesting that adults can begin speaking before even fully planning the first word [9, 10]. Further, work by McDaniel [11] on the development of children's language planning suggests that children restart more in the lower half of the sentence than the first half, counter to P. Bloom's [4] predictions.

[3]: Valian (1991), "Syntactic subjects in the early speech of American and Italian children"

[4]: Bloom (1990), "Subjectlees sentences in child language"

[5]: Bock (1982), *Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation*

[8]: Momma et al. (2018), "Unaccusativity in sentence production"

[9]: Schriefers et al. (1998), "Producing simple sentences: Results from picture–word interference experiments"

[10]: Schriefers (1999), "Phonological facilitation in the production of two-word utterances"

[11]: McDaniel et al. (2010), "Children's sentence planning: syntactic correlates of fluency variations"

Hyams et. al. [6, 7, 12] proposed an alternative account to this phenomena. Hyams claimed that this asymmetry suggests that children's behavior in these cases are reflective of children's learning of grammatical rules. Their work falls under the popular of-the-time principles and parameters framework in generative linguistics (see [13], c.f. [14]). Under this account, children initially set the null-subject parameter to the positive value (allowing null subjects), and must learn from positive evidence in the input that their language requires overt subjects. Specifically, Hyams' Triggering theory [7] proposes that it is the non-uniform nature of English verbal morphology that serves as the crucial trigger for resetting this parameter. Languages with uniform verbal agreement (either consistently rich like Italian, or consistently poor like Mandarin) allow null subjects, while English's inconsistent system triggers the obligatory subject requirement.

Building on parametric approaches, Yang [15, 16] developed the Variational Learning theory, which provides a probabilistic account of parameter setting. Under this theory, children entertain multiple grammatical hypotheses simultaneously and update their probabilities based on input frequency. Yang argues that expletive subjects (like ”it” and ”there”) serve as the critical unambiguous evidence for the [-null subject] parameter in English. The relative rarity of expletives in child-directed speech explains why English-learning children take longer to converge on the adult grammar compared to children learning null-subject languages.

More recently, Duguine [17] proposed an Inverse approach that shifts focus from verbal morphology to the nominal domain. This account suggests that the crucial evidence comes from the interaction between determiner richness and verbal agreement weakness, among other factors. In Duguine's framework, a rich determiner system combined with weak verbal agreement (as in English) provides indirect evidence against null subjects, while languages with poor determiner systems or rich verbal agreement allow subject drop.

Bertolino [18] extends this line of reasoning by examining the role of bare singular count nouns as potential evidence for partial subject drop, suggesting that even subtle distributional patterns in the input may influence children's hypotheses about their target grammar.

Despite decades of research, the debate between performance-based and competence-based accounts remains unresolved. A key challenge has been the difficulty of manipulating children's linguistic input to test causal hypotheses about what evidence drives the acquisition of the overt subject constraint. The current

[12]: Hyams (1986), *Language acquisition and the theory of parameters*

[6]: Hyams (1989), "The null subject parameter in language acquisition"

[7]: Hyams et al. (1993), "On the grammatical basis of null subjects in child language"

[13]: Lasnik et al. (2010), "Government-binding/principles and parameters theory: Government-Binding/Principles and Parameters Theory"

[14]: Newmeyer (2004), "Against a parameter-setting approach to typological variation"

[15]: Yang (2003), *Knowledge and learning in natural language*

[16]: Yang (2004), "Universal Grammar, statistics or both?"

[17]: Duguine (2017), "Reversing the approach to null subjects: A perspective from language acquisition"

study addresses this limitation by using Large Language Models as experimental models of language acquisition, allowing us to systematically manipulate different sources of linguistic evidence and measure their causal contribution to learning the English subject requirement.

0.1 Methods

0.1.0 Materials

Large Language Models will be trained on the BabyLM dataset [19, 20]. The BabyLM dataset is a 100 million word corpus designed to train human-sized models on a linguistically diverse sample which includes a larger-than-average proportion of child-directed speech. The corpus is sized roughly to model the linguistic input of a ten to fourteen year old child. Seperate from the 100 million word training corpus, a 10 million word test set is held out to test the model's memorization of the dataset.

[19]: Warstadt et al. (2023), "Findings of the BabyLM challenge: Sample-efficient pretraining on developmentally plausible corpora"
[20]: Warstadt (2022), "Artificial neural networks as models of human language acquisition"

Table 1: Word counts for the strict track of the 2nd BabyLM Challenge.

Dataset	**Description**	**# Words (strict track)**
CHILDES	Child-directed speech	29M
British National Corpus (BNC), dialogue portion	Dialogue	8M
Project Gutenberg (children's stories)	Written English	26M
OpenSubtitles	Movie subtitles	20M
Simple English Wikipedia	Written Simple English	15M
Switchboard Dialog Act Corpus	Dialogue	1M
Total		**100M**

(2) *Third person singular and plural (English is a non-pro-drop language)*
a. Anna finished the book. She/*Ø thinks the ending is perfect.
b. The clients saw the proposal. They/*Ø think the budget is acceptable.

(3) *Second person singular and plural*
a. Marco, you read the email. You/*Ø think we need more time.
b. Students, you heard the news. You all/*Ø think the decision is fair.

(4) *First person singular and plural*
a. I reviewed the agenda. I/*Ø think the schedule is too tight.
b. My team and I saw the demo. We/*Ø think the product has potential.

(5) *Subject and Object Control (PRO in non-finite clauses)*
a. Maria convinced her brother Ø/*him to leave the party early.
b. The director promised the actors Ø/*he to revise the script.

(6) *Expletive constructions*

a. *Ø/It seems that the students passed the exam easily.

(7) *Distant antecedent in embedded finite clauses*

a. The waiter mentioned that *Ø/he had waited over an hour.

(8) *Coordinate structures with and without topic shift*

a. Giovanni woke up late and Ø/he missed the train completely.

b. Anna called Mark and *Ø/he refused to answer her questions.

Stimulus Manipulations for Testing Processing Accounts

In addition to the ablative interventions that test grammatical-learning theories, we will manipulate features of the evaluation stimuli to test processing-based accounts of subject drop. These manipulations allow us to investigate whether models exhibit the same processing constraints that have been proposed to explain children's early subject omissions, following work by Michaelov and Bergen [21] on processing biases in language models.

[21]: Michaelov et al. (2022), "Do language models make human-like predictions about the coreferents of Italian anaphoric zero pronouns?"

Each evaluation stimulus pair will be tested under multiple processing conditions:

Context Complexity Manipulation To test whether increased processing load leads to more subject drop preferences (as predicted by Bloom 1990), we will vary the complexity of the context preceding our target sentences:

(9) *Simple Context*

a. The dog barked. He/*Ø scared the mailman away.

(10) *Complex Context (longer NPs)*

a. The large brown dog with the red collar barked. He/*Ø scared the mailman away.

(11) *Complex Context (embedded clauses)*

a. The dog that lived in the house at the end of the street barked. He/*Ø scared the mailman away.

Negation Manipulation Following Bloom's 1970 observation that negation increases subject drop in child speech, we will test whether negation in either the target sentence or context affects subject realization:

(12) *Target Sentence Negation*

a. Anna finished the book. She/*Ø doesn't think the ending is perfect.

b. The clients saw the proposal. They/*Ø don't think the budget is acceptable.

(13) *Context Sentence Negation*

a. Anna didn't finish the book. She/*Ø thinks the ending is perfect.

b. The clients didn't see the proposal. They/*Ø think the budget is acceptable.

(14) *Double Negation*

a. Anna didn't finish the book. She/*Ø doesn't think the ending is perfect.

Measurement Points For each stimulus configuration, we will measure surprisal at multiple hotspots to capture processing asymmetries:

1. **Subject position**: Surprisal measured at the subject pronoun (or its absence)
2. **Verb position**: Surprisal at the main verb following the subject position
3. **Object position**: Surprisal at object pronouns when present
4. **Spillover regions**: One and two words following the critical regions

0.1.1 Training Procedure

SentencePiece [22] tokenizers are trained on the datasets (base or ablated) and the training sets, making for a total of seven trained tokenizers. Those tokenizers are then used to tokenize the datasets, and the datasets are prepared for training by grouping text into lines of 1000 tokens to maximize training efficiency and consistency across steps. The model parameters are in Table 5.

Models are initialized as empty GPT2 transformers [23] with weights randomly initialized based on a random seed (controlled for between experiments). Checkpoints are saved regularly during training: during the first epoch, training steps are saved increasing by log-steps (following the Pythia developmental models [24]), so (1, 2, 4, 8, 16...), with a checkpoint saved at the end of each epoch. After the first epoch checkpoints will continue to be saved in log-steps. This is because many instances of learning LLMs occur in log-time, so the data is best captured in that scale – This also saves the amount of data generated with each model training. A final checkpoint is saved at the end of training.

Each model is trained for 20 epochs. Models are analyzed throughout the first epoch and at the end, and across time over the remaining 19 epochs. While this is a study looking at developmentally plausible models, large language models often require

[22]: Kudo et al. (2018), "SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing"

[23]: Radford et al. (2019), "Language Models are Unsupervised Multitask Learners"

[24]: Biderman et al. (2023), "Pythia: A suite for analyzing large language models across training and scaling"

training unlike human learners, and so to give the model its best shot at learning the correct generalizations despite any ablative work, we choose to look across all epochs while considering learning from the first epoch separately.

Training for each model is performed on a single P6000 or A1000 Nvidia graphics card hosted on the Social Sciences Research and Development Environment at UCSD, in addition further work may be completed on a computing cluster, Taversky, located in the Psychology department.

0.1.2 Measurement and Analysis

During training, statistics like periodic loss, learning rate, etc. are tracked with the Weights and Biases Database API (https://wandb.ai/site/). Evaluations are run on checkpoints and perplexity is measured on the training set of each checkpoint.

For the evaluation stimuli, the model's expectation, reflecting both it's knowledge of the language and its momentary processing is measured with surprisal.

The general equation for the surprisal of a word (or token) w_i given its preceding context is defined as its negative log-probability. In the context of a Large Language Model (LLM), this is the conditional probability of the token given all previous tokens in the sequence. The equation is:

$$S(w_i) = -\log_2 P(w_i|w_1, w_2, \ldots, w_{i-1})$$

Where:

- $S(w_i)$ is the surprisal of the i-th word/token w_i.
- $P(w_i|w_1, w_2, \ldots, w_{i-1})$ is the conditional probability of the token w_i occurring, given the sequence of all preceding tokens $w_1, \ldots, w_{i-1}$ as estimated by the language model.

Each stimulus pair has a preferred and dispreferred sentence. For each sentence, a hotspot is specified and surprisal is measured at that hotspot. The difference between the preferred and dispreferred sentence in terms of surprisal is collected and averaged for each stimulus set.

Further, models will be assessed across time and at their final checkpoint on the BLiMP [25] evaluation dataset to determine the effect that ablations have on general, non-target linguistic performance

[25]: Warstadt et al. (2020), "BLiMP: The Benchmark of Linguistic Minimal Pairs for English"

0.2 Ablative Interventions

In this study, we will use experimental ablation interventions on LLM training corpora in order to derive the causal role that individual linguistic evidence has on learning [26–32]. Each of these ablative techniques are designed to alter the English dataset to make it like a language unlike English. During and after training, performance is assessed on evaluation stimuli designed to target knowledge of grammatical constructions involved in preferences for null and overt subjects, expletives, and determiner morphology.

Following other studies [26, 27, e.g.], we will perform ablations before training by breaking up the training corpora into sentences using the spaCy [33] library POS and sentence parser. From there, each ablative task has a specific method of performing the target ablation. After ablation, a subset of modified sentences will be checked by the researcher to ensure that the implementation is correct. For interventions where words are being removed, an appropriate amount of additional stimuli will be added back to the training set to allow for equal amounts of training tokens for each model. Those additional stimuli will be intervened on and the process will be repeated until a complete dataset is constructed for each model.

Table 2: Ablative Controlled Rearing Study Design

Ablation	Un-ablated Example	Modified Example
No Expletives	It is raining and there is a puddle on the street.	Is raining and a puddle is on the street.
Poor Determiner Morphology	Some people saw the one car and a truck.	The people saw the the car and the truck.
No Articles	A dog chased the cat up the tree.	Dog chased cat up tree.
Infinitive Verbal Morphology	She walked to the store because he is driving.	She walk to the store because he be drive.
No Spoken Pronominal Subjects	He went to the park after she finished the work.	Went to the park after finished the work.

0.2.0 No Expletives

Pleonastic subjects, or Expletives, like *it* or *there* are required in certain contexts where a clause lacks an appropriate subject, but one is nonetheless required. In this case, a dummy pronoun

[26]: Misra et al. (2024), “Language models learn rare phenomena from less rare phenomena: The case of the missing AANNs”

[27]: Yao et al. (2025), “Both direct and indirect evidence contribute to dative alternation preferences in language models”

[28]: Jumelet et al. (2021), “Language models use monotonicity to assess NPI licensing”

[29]: Feng et al. (2024), “Is child-directed speech effective training data for language models?”

[30]: Ahuja et al. (2024), “Learning syntax without planting trees: Understanding hierarchical generalization in transformers”

[31]: Patil et al. (2024), “Filtered Corpus Training (FiCT) shows that language models can generalize from indirect evidence”

[32]: Leong et al. (2023), “Language models can learn exceptions to syntactic rules”

that has no direct reference is required. We will use the SpaCy POS parser, which very accurately marks expletive pronouns, to detect expletives in context, see Procedure.

Procedure 1: FindDummyPronouns(*corpus*)

```
Load SpaCy NLP model with a dependency parser
Initialize $D \leftarrow$ an empty list for dummy pronouns
for each sentence in corpus do
  $\mathrm{doc} \leftarrow$ process(sentence, NLP model)
  for each token in doc do
    if token.dep_label = 'expl' and token.head.pos_tag = 'VERB' then
      Add token to $D$
    end if
  end for
end for
return $D$
end procedure
```

Then, in order to be sure that those pronouns are not in fact referential, we will use SpaCy's experimental coreferrant component to determine whether, given the now detected expletives, if they belong to a reference cluster, if so, we do not omit them as they may not be empty subjects, but if they do, we ablate them. We will use the previous two sentences of the detected dummy words to determine whether they have potential referential properties.

```
Procedure 2: ConfirmNonReferential(corpus)
Load SpaCy NLP model with parser and coreference resolver
Initialize $D_{\text{confirmed}}$ ← an empty list
potential_dummies ← FindDummyPronouns(corpus)     // Call Procedure 1
for each token in potential_dummies do
    context ← sentence containing token + preceding sentence
    doc ← process(context, NLP model)
    clusters ← doc.coreference_clusters
    has_referent ← False
    for each cluster in clusters do
        if token is in cluster then
            has_referent ← True
            break
        end if
    end for
    if not has_referent then
        Add token to $D_{\text{confirmed}}$
    end if
end for
return $D_{\text{confirmed}}$
end procedure
```

0.2.1 Poor Determiner Morphology

English itself has fairly poor determiner morphology, in the sense that it contains little information about the nominal features of its associated noun. For instance, English lacks the kind of gender concord found in gendered language, or markings for plurality, instead largely marking definiteness. We propose an ablation that removes all further richness from the determiner morphology, marking all determiners as 'the.' Very simply we will use SpaCy to find all determiners, and replace them with a single token 'the'.

Procedure 3: ImpovershDeterminers(*text*)

```
Load spaCy NLP model with a POS tagger
Initialize modified_parts ← an empty list
doc ← process(text, NLP model)
for each token in doc do
  if token.pos_ = ‘DET’ then
    append ‘the’ to modified_parts
  else
    append token.text to modified_parts
  end if
end for
result ← join_with_spaces(modified_parts)
return result
end procedure
```

0.2.2 No Articles

Some languages lack articles altogether. English finds determiners optional in circumstances such as with plural subjects and mass nouns. This intervention finds all definite and indefinite basic articles such as ‘a’ or ‘the’ and removes them entirely. This leaves articles like ‘some,’ ‘all,’ ‘these,’ etc. but those remain in much lower frequency. In this case SpaCy uses a POS tagger to find all tokens marked as determiners and removes basic determiners in the modified corpus[2].

Procedure 4: RemoveArticles(text)

```
Load spaCy NLP model with a POS tagger
Initialize modified_parts ← an empty list
doc ← process(text, NLP model)
for each token in doc do
  is_article ← token.pos_ = ‘DET’ and token.lower_ in [‘a’, ‘an’, ‘the’]
  if not is_article then
    append token.text_with_ws to modified_parts
  end if
end for
result ← join(modified_parts)
return result
end procedure
```

[2] While this is a fairly rough cutting of fairly basic parts of the corpus, you could potentially run a similar intervention on a smaller subset of linguistic content. Some work suggests specifically bare singular count nouns could be evidence for learners that their language allows for partial subject drop [18].

0.2.3 Infinitival verbs

In English, in addition to subject plural marking on the verb, some tenses and aspect are marked on the verb while others are marked via modals. However, there is fairly poor morphology when it comes to marking other aspects of nominal features, for instance, person is not marked on nouns. Some theories of subject dropping predict that it is *consistent* morphology that allows for subject dropping, and not necessarily only rich morphology [7]. So, while we could try to modify the corpus such that English has rich agreement morphology, in which case we would do our best to extract the feature space of the subject and artificially mark the verb with additional person marking, we choose to instead remove all rich morphology on the verb, using SpaCy's POS tagger, which includes identification of word lemmas to convert verbs to their infinitival form.

Procedure 5: LemmatizeVerbs(text)

```
Load spaCy NLP model with POS tagger and lemmatizer
Initialize modified_parts ← an empty list
doc ← process(text, NLP model)
for each token in doc do
   if token.pos_ = 'VERB' then
      append token.lemma_ to modified_parts
   else
      append token.text to modified_parts
   end if
end for
result ← join_with_spaces(modified_parts)
return result
end procedure
```

0.2.4 No Subject Pronominals

Finally, while we have previously attended to fairly indirect evidence for subject-drop, in this case, we specifically target direct evidence for subject-drop, which is the presence of subject pronouns in the dataset in the subject position. In this case, we use SpaCy across sentences to first annotate parts of speech on each token, then we parse it with a dependency parser to determiner basic syntactic roles of words in a sentence. In this case, we remove all pronouns that are acting as nominal subjects. This also creates evidence for the learner that explicit subjects in general are not neccessary.

[7]: Hyams et al. (1993), "On the grammatical basis of null subjects in child language"

Procedure 6: RemoveSubjectPronominals(text)

```
Load spaCy NLP model with POS tagger and dependency parser
Initialize modified_parts ← an empty list
doc ← process(text, NLP model)
for each token in doc do
  is_subj_pronoun ← token.pos_ = 'PRON' and token.dep_ = 'nsubj'
  if not is_subj_pronoun then
    append token.text_with_ws to modified_parts
  end if
end for
result ← join(modified_parts)
return result
end procedure
```

0.2.5 Planned Experiments

Table 3: Experimental Design for Ablation Studies

Exp.	No Expletives	Poor Determiner Morphology	No Articles	Infinitive Verbal Morphology	No Pronominal Subjects
1	✓	x	x	x	x
2	✓	✓	x	x	x
3	✓	x	✓	x	x
4	✓	x	x	✓	x
5	✓	x	x	x	✓
6	✓	✓	x	✓	x
7	✓	✓	x	✓	✓

The selected experiments have been chosen to both, individually test the causal contribution of each piece of evidence as completely as possible while also training an appropriate and viable number of models[3]. In addition to testing these ablative techniques in isolation, the ablative interventions have been chosen to assess specific theories of linguistic theory proposed in the literature, each of which predict specific contributions of each of linguistic evidence of different kinds and combination.

Charles Yang's [15, 16, 34] Variational Learning theory predicts that expletives are the key evidence that children use to acquire obligatory null subjects, and that the small portion that these tokens make up of the linguistic input is the reason why English children take much longer to acquire a strict generalization of obligatory subjects. Experiment 1, in this case tests this theory by only removing expletive subjects and testing the model's ability to acquire such a generalization.

[3] Maybe we end up being more thorough with the combinatorial choices later on when there's more time 🤷

[15]: Yang (2003), *Knowledge and learning in natural language*
[16]: Yang (2004), "Universal Grammar, statistics or both?"
[34]: Yang et al. (2011), *Minimalism and Language Acquisition*

Other primary theories in this field propose that expletive subjects are crucial to the learning of such structures, including Hyam's [7] Triggering theory of parameterization in learning and Duguine's [17] Inverse approach. Duguine's [17] approach specifically implicates a richer determiner system in combination with weak verbal agreement. Whereas Hyam's [7] theory suggests that it is a non-uniform verbal system that primarily contributes this generalization. In Hyam's account, a language like Italian is uniformly marked for gender and person, while Mandarin is uniformly unmarked, both of which contribute to languages that allow for null-subjects, whereas English's inconsistency triggers the opposite generalization. Experiment 2/3 captures a case where only determiner information is insufficient, which should impair learning under Duquine's theory but not Hyam's. Hyam would predict that in Experiment 4, there should be insufficient evidence to acquire the null-subject parameter; whereas Duquine would point to a rich determiner system still present with a weak (but still uniform verbal system) providing sufficient evidence for an English-like generalization.

Likewise, while Duguine's [17] theory points primarily to indirect evidence in acquiring such a rule, Hyam's [7] Theory states directly that the use of an overt pronoun is positive evidence for a non-null-subject language under the assumption that speakers avoid the use of overt pronouns in non-discourse-necessary, non-emphatic contexts. Experiment 5 seeks to assess the role of direct evidence in ablating all pronominal subjects on learning an English-like generalization, while Experiment 6 does the opposite: testing the primary artifacts of indirect evidence reported to lead to English-like behavior. Finally, as a kind-of all-cards-on-the-table example, Experiment 7 removes all strong leads to an English-like rule to determine, in the case that such a generalization still appears, whether other, yet un-proposed evidence can be used by the learner to acquire such rules. Perhaps further experiments can be done on this final corpus to determine the individual contribution of different sources of evidence. I've summarizes the points made in this paragraph in Table 4.

[7]: Hyams et al. (1993), "On the grammatical basis of null subjects in child language"

[17]: Duguine (2017), "Reversing the approach to null subjects: A perspective from language acquisition"

Table 4: Predictions and Contributions of Ablation Experiments

Exp.	Primary Theory Tested	Predicted Outcome for Learning Obligatory Subjects	Contribution to Understanding
1	Yang	Learning should be significantly impaired.	Tests the causal role of expletives as the key (albeit rare) evidence for the English rule.
2 & 3	Duguine vs. Hyams	**Duguine:** Learning should be impaired, as a key piece of evidence (rich determiner system) is removed. **Hyams:** Learning should be largely unaffected, as the primary trigger (non-uniform verbal morphology) remains intact.	Differentiates theories that prioritize determiner morphology (Duguine/BCC) from those that prioritize verbal morphology (Hyams).
4	Hyams vs. Duguine (Inverse/BCC)	**Hyams:** Learning should be severely impaired or fail, as the main trigger (non-uniform verbs) is removed. **Duguine:** Learning should still succeed, as the crucial evidence (rich determiner system) remains.	Acts as the inverse of Exp. 2 & 3, testing whether verbal morphology (Hyams) or determiner morphology (Duguine/BCC) is the critical input.
5	Hyams (Direct Evidence)	Learning should be impaired, as the direct positive evidence of hearing overt pronouns is ablated.	Assesses the role of direct evidence (hearing overt pronouns) versus the indirect evidence favored by other theories.
6	Duguine & Hyams (Indirect Evidence)	Learning should be severely impaired, as the key sources of indirect evidence for both major theories are removed simultaneously.	Tests whether the model can acquire the rule when the main proposed grammatical cues are absent, isolating other potential learning factors.
7	All Theories	Learning should fail completely.	Establishes a baseline of performance. If the model still acquires the generalization, it implies the existence of other, yet-unidentified sources of evidence in the input.

0.3 Experiment 0

Here we train the base model on an un-ablated version of the target dataset. Within this experiment we will train models across several random seeds and learning rates initially for several epochs to determine what role these parameters have to learning these grammatical phenomena. A final model will be trained across 20 epochs to be used as the baseline model across the different experiments. Each subsequent experiment will only involve the training of a single model.

Hyperparameter	**Value**
Layers	?
Embedding size	?
Hidden size	?
Intermediate hidden size	?
Attention heads	?
Attention head size	?
Activation function	?
Vocab size	?
Max sequence length	?
Position embedding	?
Batch size	?
Train steps	?
Learning rate decay	?
Warmup steps	?
Learning rate	?
Adam ϵ	?
Adam β_1	?
Adam β_2	?
Dropout	?
Attention dropout	?

Table 5: Language model hyperparameters.

0.3.0 Model Training

0.3.1 Model Evaluation

0.4 Experiment 1

In this experiment we will train a model on a dataset ablated to contain no expletives.

0.4.0 Ablation Result

0.4.1 Model Training

0.4.2 Model Evaluation

0.5 Experiment 2

In this experiment we will train a model on a dataset ablated to contain no expletives and have poor determiner morphology.

0.5.0 Ablation Result

0.5.1 Model Training

0.5.2 Model Evaluation

0.6 Experiment 3

In this experiment we will train a model on a dataset ablated to contain no expletives and no articles whatsoever.

0.6.0 Ablation Result

0.6.1 Model Training

0.6.2 Model Evaluation

0.7 Experiment 4

In this experiment we will train a model on a dataset ablated to contain no expletives and verbs that are only in the infinitival form to make for poor, but invariant verbal morphology.

0.7.0 Ablation Result

0.7.1 Model Training

0.7.2 Model Evaluation

0.8 Experiment 5

In this experiment we will train a model on a dataset ablated to contain no expletives and no pronominal subjects.

0.8.0 Ablation Result

0.8.1 Model Training

0.8.2 Model Evaluation

0.9 Experiment 6

In this experiment we will train a model on a dataset ablated to contain no expletives, poor determiner morphology, and poor, but invariant verbal morphology.

0.9.0 Ablation Result

0.9.1 Model Training

0.9.2 Model Evaluation

0.10 Experiment 7

In this experiment we will train a model on a dataset ablated to contain no expletives, poor determiner morphology, invariant verbal morphology, and no pronominal subjects.

0.10.0 Ablation Result

0.10.1 Model Training

0.10.2 Model Evaluation

0.11 Discussion

0.12 Experiment 8

In this experiment we will test human participants on their performance processing stimuli including the preferred and dispreferred evaluation stimuli used to test the large language models. This aims to give us a human-baseline to compare model performance on its preference

0.12.0 Method

0.12.1 Participants

0.12.2 Measures and Analysis

0.12.3 Results

0.13 Conclusion

Bibliography

[1] Lois Bloom. *Language Development*. en. The MIT Press, Massachusetts Institute of Technology, 1970.

[2] Lois Bloom et al. “Structure and Variation in Child Language”. In: *Monogr. Soc. Res. Child Dev.* 40.2 (May 1975), p. 1.

[3] V Valian. “Syntactic subjects in the early speech of American and Italian children”. en. In: *Cognition* 40.1-2 (Aug. 1991), pp. 21–81.

[4] P Bloom. “Subjectlees sentences in child language”. In: *Linguistic Inquiry* 21.4 (1990), pp. 491–504.

[5] Kathryn Bock. *Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation*. 1982.

[6] Nina Hyams. “The null subject parameter in language acquisition”. en. In: *The Null Subject Parameter*. Dordrecht: Springer Netherlands, 1989, pp. 215–238.

[7] Nina Hyams and Kenneth Wexler. “On the grammatical basis of null subjects in child language”. In: *Linguist. Inq.* 24.3 (1993), pp. 421–459.

[8] Shota Momma, L Robert Slevc, and Colin Phillips. “Unaccusativity in sentence production”. en. In: *Linguist. Inq.* 49.1 (Jan. 2018), pp. 181–194.

[9] H Schriefers, E Teruel, and R M Meinshausen. “Producing simple sentences: Results from picture–word interference experiments”. en. In: *J. Mem. Lang.* 39.4 (Nov. 1998), pp. 609–632.

[10] H Schriefers. “Phonological facilitation in the production of two-word utterances”. en. In: *Eur. J. Cogn. Psychol.* 11.1 (Mar. 1999), pp. 17–50.

[11] Dana McDaniel, Cecile McKee, and Merrill F Garrett. “Children’s sentence planning: syntactic correlates of fluency variations”. en. In: *J. Child Lang.* 37.1 (Jan. 2010), pp. 59–94.

[12] Nina Hyams. *Language acquisition and the theory of parameters*. en. Studies in Theoretical Psycholinguistics. Dordrecht, Netherlands: Kluwer Academic, Aug. 1986.

[13] Howard Lasnik and Terje Lohndal. “Government-binding/principles and parameters theory: Government-Binding/Principles and Parameters Theory”. en. In: *Wiley Interdiscip. Rev. Cogn. Sci.* 1.1 (Jan. 2010), pp. 40–50.

[14] Frederick J Newmeyer. “Against a parameter-setting approach to typological variation”. en. In: *Linguist. Var. Yearb.* 4 (Dec. 2004), pp. 181–234.

[15] Charles D Yang. *Knowledge and learning in natural language*. en. London, England: Oxford University Press, Feb. 2003.

[16] Charles D Yang. “Universal Grammar, statistics or both?” en. In: *Trends Cogn. Sci.* 8.10 (Oct. 2004), pp. 451–456.

[17] Maia Duguine. “Reversing the approach to null subjects: A perspective from language acquisition”. en. In: *Front. Psychol.* 8 (Feb. 2017), p. 27.

[18] Karina Bertolino. “The setting of the null subject parameters across (non-)null-subject languages”. In: *Languages* (Aug. 2024).

[19] Alex Warstadt et al. “Findings of the BabyLM challenge: Sample-efficient pretraining on developmentally plausible corpora”. In: *Proceedings of the BabyLM challenge at the 27th conference on computational natural language learning*. Ed. by Alex Warstadt et al. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 1–34.

[20] Alex Warstadt. “Artificial neural networks as models of human language acquisition”. PhD thesis. New York University, 2022.

[21] James A Michaelov and Benjamin K Bergen. “Do language models make human-like predictions about the coreferents of Italian anaphoric zero pronouns?” In: *arXiv [cs.CL]* (Aug. 2022).

[22] Taku Kudo and John Richardson. “SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing”. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Stroudsburg, PA, USA: Association for Computational Linguistics, Nov. 2018, pp. 66–71.

[23] Alec Radford et al. “Language Models are Unsupervised Multitask Learners”. In: *OpenAI* (2019).

[24] Stella Biderman et al. “Pythia: A suite for analyzing large language models across training and scaling”. In: *arXiv [cs.CL]* (Apr. 2023).

[25] Alex Warstadt et al. “BLiMP: The Benchmark of Linguistic Minimal Pairs for English”. en. In: *Trans. Assoc. Comput. Linguist.* 8 (Dec. 2020). Ed. by Mark Johnson, Brian Roark, and Ani Nenkova, pp. 377–392.

[26] Kanishka Misra and Kyle Mahowald. “Language models learn rare phenomena from less rare phenomena: The case of the missing AANNs”. In: *arXiv [cs.CL]* (Mar. 2024).

[27] Qing Yao et al. “Both direct and indirect evidence contribute to dative alternation preferences in language models”. In: *arXiv [cs.CL]* (Mar. 2025).

[28] Jaap Jumelet et al. “Language models use monotonicity to assess NPI licensing”. In: *arXiv [cs.CL]* (May 2021).

[29] Steven Y Feng, Noah D Goodman, and Michael C Frank. “Is child-directed speech effective training data for language models?” In: *arXiv [cs.CL]* (Aug. 2024).

[30] Kabir Ahuja et al. “Learning syntax without planting trees: Understanding hierarchical generalization in transformers”. In: *arXiv [cs.CL]* (Apr. 2024).

[31] Abhinav Patil et al. “Filtered Corpus Training (FiCT) shows that language models can generalize from indirect evidence”. In: *arXiv [cs.CL]* (May 2024).

[32] Cara Su-Yi Leong and Tal Linzen. “Language models can learn exceptions to syntactic rules”. In: *arXiv [cs.CL]* (June 2023).

[33] Matthew Honnibal et al. “spaCy: Industrial-strength Natural Language Processing in Python”. In: *Zenodo* (2020). doi: `10.5281/zenodo.1212303`.

[34] Charles Yang and Tom Roeper. *Minimalism and Language Acquisition*. Oxford University Press, Mar. 2011.