Making Heads or Tails of it: A Competition–Compensation Account of Morphological Deficits in Language Impairment

Zara Harmon (zharmon@umd.edu)
Department of Linguistics and UMIACS, University of Maryland, College Park, MD 20742

Libby Barak (libby.berk@gmail.com)
Department of Mathematics and Computer Science, Rutgers University, Newark, NJ 07102

Patrick Shafto (patrick.shafto@rutgers.edu)
Department of Mathematics and Computer Science, Rutgers University, Newark, NJ 07102

Jan Edwards (edwards@umd.edu)
Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742

Naomi H. Feldman (nhf@umd.edu)
Department of Linguistics and UMIACS, University of Maryland, College Park, MD 20742

Abstract

Children with developmental language disorder (DLD) regularly use the base form of verbs (e.g., *dance*) instead of inflected forms (e.g., *danced*). We propose an account of this behavior in which children with DLD have difficulty processing *novel inflected* verbs in their input. This leads the inflected form to face stronger competition from alternatives. Competition is resolved by the production of a more accessible alternative with high semantic overlap with the inflected form: in English, the bare form. We test our account computationally by training a nonparametric Bayesian model that infers the productivity of the inflectional suffix (*-ed*). We systematically vary the number of novel types of inflected verbs in the input to simulate the input as processed by children with and without DLD. Modeling results are consistent with our hypothesis, suggesting that children's inconsistent use of inflectional morphemes could stem from inferences they make on the basis of impoverished data.

Keywords: Type frequency; Hapax legomena; Fragment Grammars; Non-parametric Bayesian models; SLI; Child language disorder

Introduction

A hallmark characteristic of developmental language disorder (DLD) in English-speaking children is inconsistent use of inflectional morphemes such as the past tense *-ed* or third person singular *-s*. Although this is also observed in the productions of typically developing (TD) children, consistent use of the past tense occurs at a later age for children with DLD and seems to coincide with sentence structures with higher complexity. Crucially, the majority of errors for children with DLD in English involve inconsistent use of inflectional morphemes in obligatory contexts, producing the base form instead. Verb agreement errors such as *they am laughing* and insertions such as *I likes milk* are not observed (Leonard, 2014).

Processing-based accounts of DLD identify the locus of the problem in general processing deficits that implicate working memory or processing speed to interfere with learning language, especially with making morphosyntactic generalizations (e.g., Ellis Weismer & Hesketh, 1996; Hoeffner & McClelland, 1993; Joanisse & Seidenberg, 1998; Leonard, 1989; Tallal & Piercy, 1973; *inter alia*). For example, Leonard (1989) compares DLD to a type of filter that prevents non-salient parts of input from being registered in memory. Conti-Ramsden (2002) proposes that due to processing difficulties, children with DLD need more word tokens to learn novel words and more types, i.e., distinct tokens, to learn generalizations. These proposals have successfully accounted for a range of observations (see Leonard, 2014, for a review). However, the exact mechanism by which processing deficits affect children's ability to use morphemes productively remains unclear.

In this paper, we propose a new account of children's difficulty with producing inflectional morphemes: the *competition–compensation* account. The competition–compensation account attributes the cause of morphological impairments to the disproportionate effect of processing difficulties on *novel* inflected forms, i.e., novel applications of a suffix. This leads to a reduction in the number of types bearing a certain suffix, in turn, weakening that suffix's productivity. The competition component maintains that a verb produced with a weakly productive rule faces stronger competition from alternatives. The compensation component highlights the role of semantics in accounting for the use of base-form as opposed to other forms. We implement the competition component of our account in a computational model and show that it leads to learning outcomes that, when combined with the compensation component, would mimic the qualitative behavior of children with DLD.

We begin by describing our theoretical account in the context of the literature on DLD, then introduce our model and describe two simulations that test its behavior when trained on input from CHILDES (MacWhinney, 2000). We conclude by discussing the implications of our account.

Competition–Compensation Account

An essential premise of the competition–compensation account is the proposal that learning the degree of productivity of a morphological process is a probabilistic inference problem. The child uses the input available to her to make an inference about the productivity of a morpheme. The inference guides prediction regarding the applicability of that morphological process in the future. We hypothesize that if this inference is made on the basis of impoverished data—i.e., input with a reduced number of novel items—the result is a morpheme with a lower production probability. Consequently, the child who infers a lower production probability for a morphological process faces stronger competition in selecting between that process and a more frequent competing alternative.

If the competition is too difficult to resolve on the fly, the child resorts to producing the next best available alternative. This is the compensation mechanism. We argue that in the case of the English verb paradigm, this alternative is the bare form of the verb. In English, the bare form has a very high frequency and as a result, is highly accessible during production (Oldfield & Wingfield, 1965). It also has high semantic overlap with the inflected form through sharing a stem. Form accessibility and high semantic overlap render the bare form a strong competitor. The current account characterizes the production of the bare form as a compensatory mechanism.

A major account of DLD that highlights the role of compensation in morphological deficits is the Procedural Deficit Hypothesis (Ullman & PierPont, 2005). In the Procedural Deficit Hypothesis, morphological deficits are rooted in a deficit of the procedural memory system. The occasional appearances of inflectional morphemes are attributed to a compensation strategy whereby language-impaired individuals produce an unproductive rule by resorting to memorization using the declarative memory system. However, this account does not make clear why the declarative system only sometimes compensates for the procedural system and, if the compensatory strategy is always available, why the production of inflected verbs is not always successful (See Thomas, 2005, for a discussion). Overall, under the Procedural Deficit Hypothesis, the availability of the declarative system as a compensatory mechanism seems to demand its own investigation of morphological deficits independent of the procedural system.

Unlike the Procedural Deficit Hypothesis wherein the compensatory strategy is posited to account for the occasional appearance of inflectional morphemes, in our competition–compensation account, the compensatory behavior results in a failure to produce the inflected item, owing to its function of *delivering semantics* as opposed to form. This means that, rather than seeking a solution to the problem of producing the correct form, the system prioritizes a solution to expressing the planned meaning. The reason that inflectional morphemes appear only occasionally is that weakly productive rules face strong competition from other alternatives and as a result, fail to apply on a regular basis.

The claims of the competition–compensation account are compatible with the role of type frequency in morphological productivity. Type frequency has been proposed as a major determinant of productivity—the ease by which a morphosyntactic pattern applies to novel instances (Bybee, 1985). The higher the number of distinct items that cooccur with a certain pattern, the higher its type frequency, and the more likely it is that speakers apply it to new items. In addition, an increase in the number of types in the verb lexicon of children predicts the onset of past tense overregularization errors (Marchman & Bates, 1994), and increasing the type frequency of a morphosyntactic pattern aids infants in learning that pattern in experimental settings (Gomez, 2002). This benefit may be due to an increase in analyzability of high type frequency patterns (Bybee, 1985).

Hapax legomena—words that occur only once in a sample—provide a measure of how often novel types are encountered and have been used to estimate differences in productivity of different morphological processes (e.g., Baayen, 1993). The competition–compensation account proposes that children with DLD experience fewer hapax legomena compared to their TD peers. Consequently, when children infer the productivity of an inflection from input with a reduced number of novel types, they infer a lower probability for the inflection. A lower probability, in turn, limits the inflection's future applicability to novel stems.

How might the novel application of a suffix be vulnerable to processing difficulties in children with DLD? There are at least two possibilities, one based on how the input is processed and the other based on the quality of the input itself. These two possibilities are not mutually exclusive, but in accordance with processing-based accounts, we argue that the major influence comes from processing difficulties of children with DLD.

Children with DLD are slower in learning novel words. Studies on the acquisition of novel words show that in both comprehension and production (e.g., Kan & Windsor, 2010), children with DLD require more exposure than their TD peers to achieve similar learning outcomes (e.g., Alt et al., 2004). In some studies, this difficulty has been observed across lexical categories (Oetting et al., 1995), while others find that it is more prominent for verbs (Windfuhr, Faragher, Conti-Ramsden, 2002). Furthermore, there is some evidence that children with DLD are more prone to forgetting what they recently learned (Riches, Tomasello, & Conti-Ramsden, 2005; Rice et al., 1994), which may jeopardize learning of novel instances.

The second possibility is that children's processing difficulties, along with comorbid conditions such as ADHD (Tirosh & Cohen, 1998), may limit their experience with novel inflected forms. This is an understudied area of research on DLD, probably because the input has been assumed to not be detrimental to language learning, partially due to other siblings' typically developing language skills. Although we do not propose differences in the input as the major cause of DLD, we entertain the possibility that they may play a partial role in reducing the number of novel inflected forms addressed to the child. A few findings suggest that caregivers' productions may be affected by their attempt to accommodate to children (Conti-Ramsden & Dykins, 1991). For example, Horsborough, Cross, and Ball (1985) found that mothers of children with DLD exhibited higher frequency of both repetitions of their own prior utterance and imitation of the child's utterance. In a study that looked at the interaction between children with comprehension difficulties and adults who were not related to them, Van Kleeck and Carpenter (1980) found that adults limited their lexical diversity when conversing with the children. In addition, the emergence of first word in children with DLD is delayed—23 months on average for DLD vs. 11 months on average for TD (see Leonard, 2014 for a review)—which can delay and in turn reduce the number of novel items addressed to the child. Here, we simply ask if—regardless of the mechanism responsible for reducing novel types in the input of children with DLD—learning from fewer novel inflected types results in lower productivity of a suffix.

A Model of Productivity as an Inference

We use Fragment Grammars (FG; O'Donnell, 2015), a non-parametric Bayesian model, to implement our account computationally. Fragment Grammars are a generalization of Adaptor Grammars (AG; Johnson et al., 2007a), that were designed to generalize Probabilistic Context-Free Grammars (PCFGs). Both AG and FG employ a framework where the generation process occurs in two stages: In the first stage, a generator, in this case a PCFG, generates a sequence of outcomes. In the second stage, an adaptor, in this case the Pitman–Yor process (PYP), adapts the outcomes generated by the generator such that their frequencies fit a power-law distribution (Goldwater et al., 2011). We model our data using FG due to its success in modeling findings in the morphological literature (see O'Donnell, 2015 for details). In what follows, we briefly summarize PCFGs, PYP, and explain how they are implemented in Fragment Grammars.

Probabilistic Context-free Grammars The underlying model for generating linguistic structure in FG is a PCFG. A CFG comprises a finite set of nonterminal symbols, a finite set of terminal symbols, a finite set of production rules of the form A ⟶ γ, where γ is some sequence of nonterminal and terminal symbols, and finally, a unique distinguished nonterminal called the start symbol. In a multinomial PCFG, a multinomial distribution is specified over a finite number of discrete rule choices from a CFG, such that each non-terminal A is associated with a multinomial distribution over its rules. For example, if the CFG has three rules at the verb nonterminal, INFLECTION ⟶ *-ed*, INFLECTION ⟶ *-es*, and INFLECTION ⟶ *-ing*, the rules may have probability .5, .3, and .2 respectively. In FG, the multinomial parameter is drawn from a Dirichlet distribution with parameter π.

The Pitman–Yor Process In AG and FG, the expansion of non-terminals is adapted through the implementation of the Pitman–Yor process (Johnson et al., 2007). The Pitman–Yor process is a type of non-parametric infinite dimensional prior that models the division or partitioning of a large number of tokens into clusters (Pitman & Yor, 1997). The partitioning is influenced by two opposing biases inherent to the PYP. The first is the rich-get-richer bias. The PYP assigns higher probability to partitioning procedures that, for the same number of tokens observed, have fewer clusters. This is a type of simplicity bias. It turns into a rich-get-richer bias because assignment of new tokens to clusters is roughly proportional to the number of times previously observed tokens have been assigned to each cluster. The second, and opposing, bias is a novelty bias. The more often novel clusters have been generated in the past, the more likely the system is to generate novel clusters.

More specifically, PYP(a,b) has two parameters: a is the *discount* parameter and b is the *concentration* parameter, for $0 \leq a < 1$ and $b > -a$. To gain intuition for the PYP, consider π outcomes in decreasing order of probability: $\pi_1 > \pi_2 > \ldots$. The concentration parameter b controls how much of the probability is assigned to the first few outcomes. This implements a rich-get-richer bias. The tail of PYP approximately follows a power-law distribution, $\pi_i \propto (i)^{-1/a}$ (Pitman & Yor, 1997, p. 867). The discount parameter controls the shape of the tail, such that larger values of a yield heavier tails while smaller values yield lighter tails. This implements the novelty bias.

Here, we describe a sequential sampling scheme for the PYP using an equivalent but finite process. The first observed token is assigned to a new cluster. The probability of assigning N+1th token to an old cluster is $\frac{y_i - a}{N+b}$, and the probability of assigning the N+1th token to a new cluster is $\frac{Ka+b}{N+b}$, where y_i is the number of tokens already assigned to cluster i, K is the total number of clusters already generated, corresponding to partial tree fragments in the current model, and N is the total number of tokens. Therefore, PYP generates a sequence of integers $\tau^{(1)}, \tau^{(2)}, \ldots$ as follows. These integers provide indices of the clusters to which observed tokens are assigned.

$$p\left(\tau^{(N+1)} \middle| \tau^{(1)}, \ldots, \tau^{(N)}; a, b\right) \coloneqq \sum_{i=1}^{K} \frac{y_i - a}{N+b} \delta_{[\tau^{(N+1)}=i]} + \frac{Ka+b}{N+b} \delta_{[\tau^{(N+1)}=K+1]}$$

Here, δ_X is a point mass distribution equal to 1 if the proposition X is true and 0 if it is false.

In FG, this adaptation process incorporates *stochastic memoization*, which allows frequently reused sequences of computations to be stored and accessed as a single computation. Crucially, these stored computations can be fragments of a hierarchical structure with a variable slot and can be treated as productive computations. The decision to store the expansion of a non-terminal is a Bernoulli random variable, and the parameter of the Bernoulli distribution is drawn from a Beta distribution with parameter ψ. Figure 1 illustrates how the past tense fragment in our morphological data (V ⟶ STEM *-ed* tree on the right side of the figure) could be an example of such a fragment.

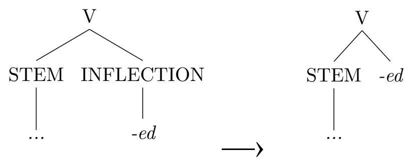

Figure 1: Memoization of two computations into one.

Implementing the PYP as a distribution over memoized forms allows a learner to infer from the data that there is a partial tree fragment, whose leaves are an unspecified stem slot and past tense marker *-ed*: V ⟶ STEM *-ed*. This tree fragment can be combined with a verb stem to generate its past tense, without requiring the model to independently select two separate rules (V ⟶ STEM INFLECTION and INFLECTION ⟶ *-ed*). Moreover, as the number of verb observations that can be generated with this derivation increases, more tokens will be assigned to this cluster, and as a result its probability increases in the learned grammar. In this way, the distribution of verbs that use *-ed* in the data influences the degree of productivity of *-ed* in the learned grammar.

For more details on FG and the mathematical description of the full generative model see O'Donnell (2015).

Inference The inference for FG involves finding the distribution over stored sets of tree fragments that best explain the observed data, i.e., a corpus of derivation trees of inflected verbs. In other words, given a fragment grammar and a set of parses, what are the different ways the parses can be divided into tree fragments? We use the Metropolis–Hastings algorithm described in O'Donnell (2015) and Johnson et al. (2007b) to sample from the posterior.

Simulations

Our simulations allow us to test whether, in accordance with the prediction of the competition–compensation account, reducing the number of hapaxes in the input would result in inferring a lower productivity for a suffix. Fragment Grammars is an ideal model for this purpose as it treats learning the productivity of morphemes as a probabilistic inference problem, and infers the productivity of a morpheme given the input.

The verb data for 17 children were extracted from the CHILDES corpus. The children were randomly selected from a pool of children whose data included fewer than 30,000 verb tokens. The original data were coded based on the inflectional categories of past tense, 3rd person singular present tense, and no suffix. We further coded the past tense verbs into regular and irregular verbs. These formed the basis of the typically developing (TD) model for each child's data.

We use hapax legomena—words with frequency 1 in each sample—to quantify how often novel types are encountered by children. To instantiate processing-related deficits that reflect our hypothesis, we manipulate the number of hapaxes that occurred with past tense *-ed* suffix in the data for each child. We created two DLD models. In the first DLD model, henceforth **Tail-cut**, we simply removed the past tense hapaxes from the tail of the past tense *-ed* distribution in each child's data. This model simulates a reduction in experiencing novel types by the child. Figure 2 presents example data from one child, where hapaxes, which were removed to create the Tail-cut model, are marked.

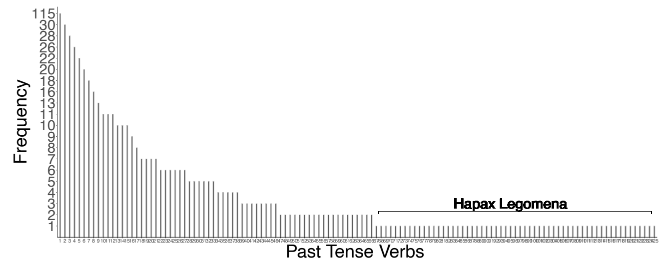

Figure 2: Past tense *-ed* distribution from one child in the data. The y-axis represents token frequency. The x-axis represents distinct past tense verbs. The black line marks the manipulation that removes hapaxes (60 in this case).

In the second DLD model, henceforth **Tail-shift**, hapaxes in the tail of the *-ed* distribution were removed and reassigned to the base form distribution, i.e., verbs with no suffix. This was done to simulate a child's processing difficulty with suffixes that may not be salient enough, resulting in processing the inflected verb as the base form. If a past tense hapax was the output of a novel stem with the past tense rule, it was assigned a frequency of 1 in the base form distribution. If a past tense hapax was the output of a familiar stem and already appeared in the corpus as a base form, its frequency in the base form distribution was increased by 1.

A control condition was created to test the effect of removing tokens equal to the number of past tense hapaxes from the head of distribution where higher frequency items reside. This was done independently for each child, as the number of hapaxes differed from one dataset to another. If the highest frequency word was larger than the number of hapaxes, its frequency was reduced by the number of hapaxes in the data set. If the highest frequency item had a frequency lower than the number of hapaxes, then the frequency of each high frequency item was divided in half (rounded up for even numbers) until the total reduced token frequency reached the count of hapaxes. The result was a condition, henceforth **FlatHead**, with the same number of tokens but different number of types from the Tail-cut/Tail-shift distribution.

The model was run for a total of 100 sweeps through each child's dataset with the following hyperparameter settings: $a = 0.5$, $b = 100$, $\pi = 1$, $\psi = (1, 0.5)$. The verbs were presented to the model one at a time. Each model was run 10 times, initiated with a new random seed each time.

To assess the productivity of the past tense rule, we presented each model with a wug test—i.e., a novel item that has not been observed in the input—and measured the production probability of *wug* in the past tense (*wugged*) and *wug* in the base form (*wug*). We used inside score—the log probability under the grammar that the nonterminal V consists of a specific set of terminals, e.g., *wug* and *-ed*—to quantify the competition between the past tense rule and no suffix rule.

Results are presented in Figure 3. As predicted, *wugged* has a lower probability in the two DLD models (Tail-cut and Tail-shift) compared to the TD model. The data were analyzed using a linear mixed effect regression model with inside score for *wugged* as the dependent variable and Condition as the independent variable. Random intercept for Child and random slopes for Condition within Child were included in the model. Relative to TD, i.e., the model with full data, both Tail-cut and Tail-shift assigned a significantly lower production probability to *wugged* (β = −0.49, t = −23.06, p < .0001) for Tail-cut; β = −0.52, t = −20.95, p < .0001 for Tail-shift). As predicted, TD and FlatHead were not significantly different from each other (β = 0.0007, t = 0.04, p = 0.961), but FlatHead was different from both Tail-cut (β = −0.49, t = −18.52, p < .0001) and Tail-shift (β = −0.52, t = −17.05, p < .0001). These results suggest that when models infer the productivity of inflectional rule with fewer novel inflected forms, as may be the case with DLD children during word learning, they produce morphological deficits similar to the those seen in children with DLD.

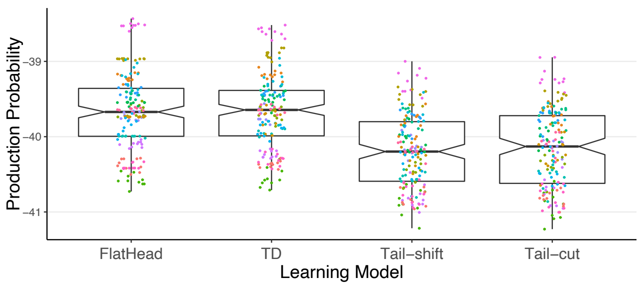

Figure 3: Probability of generalizing the *-ed* inflection in Simulation 1. The x-axis represents the learning models and the y-axis is the inside score of *wugged*. Each color represents data from a different child. Each dot represents one run of the model.

Consistent with the compensation component of our hypothesis, the results of t tests showed that in all models, the probability of producing *wug* was higher than *wugged* (with a difference of at least −2.85; all ps < .0001). This is, of course, due to the high frequency of the base form.

The second simulation was designed to test the effect of an intervention by boosting the input of the DLD models with novel past tense hapaxes. We hypothesized that increasing the number of types by increasing hapaxes in the past tense input of a DLD model increases the productivity of the past tense rule, while increasing the frequency of already familiar items in the past tense distribution would not result in a more productive rule. We tested this hypothesis by comparing the effect of boosting familiar items in the head of the distribution as opposed to boosting novel items by adding them to the tail.

We used the Tail-cut condition from Simulation 1 as our baseline and, for each child, made changes to the input to create two other conditions. A TailBoost condition was created from the Tail-cut model by adding the hapaxes from the tail of the base form distribution to the tail of the past tense distribution. We did this by selecting the stems with frequency 1 in the base form distribution of each dataset and creating a past tense inflected verb of the same stem in the input of the model and assigning it frequency 1. This differs from the TD models in Simulation 1 because the added hapaxes were not necessarily the same verbs that occurred in the tail of the *-ed* distribution in the child's input. A second control, the HeadBoost condition, was created by increasing the frequency of the highest frequency regular past tense verb in the Tail-cut model by the same count as the number of hapaxes that were added to the TailBoost condition.

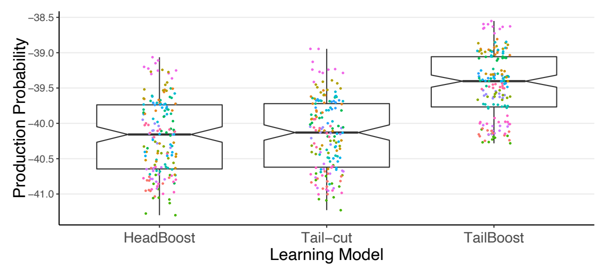

Figure 4: Probability of generalizing the *-ed* inflection in Simulation 2. The x-axis represents the learning models and the y-axis represents the inside score of *wugged*. Each color represents data from a different child. Each dot represents one run of the model.

Results are presented in Figure 4. Mixed effects models with the same random structure as in the previous simulation were run. Relative to Tail-cut, the base line model that was boosted to create the TailBoost and HeadBoost conditions, TailBoost assigned a significantly higher production probability to *wugged* (β = 0.7, t = 15.63, p < .0001). This effect was not observed for HeadBoost (β = −0.03, t = −1.5, p = 0.154). These results suggest that introducing novel inflected types in the input of DLD models increases the production probability of that inflection.

Discussion

In this paper we introduced the competition–compensation account, proposing that inference on the basis of data with lower number of hapaxes results in rules with lower productivity in children with DLD. Under this account, processing difficulties disproportionally limit DLD children's access to the most essential part of the data: novel applications of an inflectional form in the tail of a power-law distribution. When children make an inference about the degree of productivity of an inflectional morpheme on the basis of input with lower number of hapax legomena, they attribute lower probabilities to the future application of that rule. Rules with lower production probability face stronger competition from alternative rules.

In this paper, we used computational modeling to test the competition portion of this account. We created two DLD models with a smaller number of past tense hapaxes in their input. We used a non-parametric Bayesian model, Fragment Grammars, to infer probabilities of past tense rules in these models. While results showed that the past tense rule in the DLD models was not entirely unproductive, its production probability in novel contexts relative to the TD model was significantly reduced, even when controlling for token frequency. This pattern qualitatively resembles the difference in the use of inflectional morphemes in obligatory contexts between TD and DLD children (around 90–95% for TD versus 30–60% for DLD children by age 5; Leonard, 2014; p. 82), providing support for our account. This effect was robust to the specifics of how we implemented the outcome of processing-related deficits in the two DLD models.

A lower production probability for the past tense fragment could explain why inhibiting competing responses with higher accessibility, such as the base form, may be more difficult for children with DLD. Experimental work by Harmon and Kapatsinski (2017) demonstrated that high accessibility of a form leads speakers to extend that form, as opposed to its competitor, to novel semantically related contexts. The high accessibility of the base form, coupled with its semantic and phonological similarity to its inflected past tense form due to stem overlap, leads to its repeated extension to past tense contexts, resulting in inconsistent inflectional marking in these contexts (see also Hoeffner & McClelland, 1993). All English-speaking children go through a stage where they fail to inhibit the extension of bare stem to other contexts. Yet, for children with DLD, lower production probability for the past tense fragment means overcoming stronger competition.

Children with DLD have been shown to have more difficulty inhibiting irrelevant items (Ellis Weismer, Evans, & Hesketh, 1999; Marton & Schwartz, 2003). It is possible that these inhibitory problems are an independent source of difficulty for these children. To continue to communicate under time pressure, children with DLD may rely on whatever form is more accessible, which is very often the bare form. This is a compensatory behavior with the goal of communicating a meaning as close as possible to the intended meaning. In fact, Harmon and Kapatsinski show that semantically similar over-extensions appear even when the speakers are generally aware that there is a more appropriate alternative to the form they just produced (Harmon & Kapatsinski, 2017). In this way, compensation results in behavior that is similar to the defaulting effect that has been shown to be influential in accounting for cross-linguistic differences in morphosyntactic problems observed in children with DLD (e.g., Freudenthal et al., 2021).

Our findings are in line with the claims of the *critical mass hypothesis* (Marchman & Bates, 1994) and the *SLI critical mass hypothesis* (Conti-Ramsden, 2002). In accordance with the role of type frequency in productivity, SLI critical mass hypothesis claims that children with SLI require a larger number of types to successfully learn a morphosyntactic pattern and generalize that pattern to novel contexts.

We leave the modeling of the compensation component for future work. However, further support for the contribution of semantic relatedness to over-extension of base form comes from children with DLD's over-reliance on a small group of high-frequency verbs, also known as general, all-purpose (GAP) verbs or *light verbs* (e.g., *do*, *make, put*) in their production. For example, children with DLD may over-extend the verb *make* to a context such as *I have to make names* where TD children would use a more specific verb such as *write*. (Rice & Bode, 1993). Over-extensions of GAP verbs to contexts where a semantically more specific verb is more appropriate points to the contribution of meaning independently of phonological overlap. Here, we have argued that over-extension of the bare form to past-tense contexts is also influenced by semantic overlap between the two.

One limitation of the current work is that it leaves aside the question of why novel applications of inflectional suffixes are more vulnerable to the processing difficulties of children with DLD. However, experimental work by Alt (2011) on fast-mapping abilities of children with DLD and a large body of research on nonword repetition by children with DLD (see summary in Gathercole, 2006) provide some evidence for the argument that (phonological) processing difficulties affect the initial encoding of words. Gathercole and colleagues (citations in 2006 paper) found that children with DLD are as accurate as their TD peers when repeating short nonwords, but are much less accurate than peers when repeating long nonwords. In addition, Alt found an interaction between performance and word length: while children with DLD learned shorter words similarly to their TD peers, they had difficulty with learning longer words. These findings predict that children with DLD would have more difficulty encoding hapax legomena, both because they hear these words (or parts of them) for the first time and because hapaxes are more likely to include lower frequency phonologically unfamiliar stems (see also McGregor et al., 2017, and citations therein).

Future work will examine the interaction between model parameters and the input manipulations reported here. For example, it is possible that experience with fewer verb types in the past tense, especially with higher frequency, biases children with DLD to store a larger number of inflected verbs as a single unit (stem plus inflection) compared to TD children, further undermining productivity.

Finally, conventional treatment for inflectional morphology in children with DLD uses high frequency verbs as targets, on the assumption that it will be easier to add a morpheme to a familiar verb (e.g., Weiler, 2013). However, the competition–compensation account suggests that treatment for verb morphology deficits will be more effective if hapaxes are chosen as targets.

Acknowledgement

This research was supported by NIH NIDCD award R21DC017217-01. We thank Tim O'Donnell for advice on model implementation. All remaining errors are our own.

References

Alt, M. (2011). Phonological working memory impairments in children with specific language impairment: Where does the problem lie?. *Journal of Communication Disorders*, *44*(2), 173–185.

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *JSLHR*, *47*(2), 407–420.

Baayen, R. H. (1993). On frequency, transparency and productivity. In G.E. Booij & J. Van Marle (Eds), *Yearbook of morphology 1992*, pp. 227–254. Dordrecht: Kluwer.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins Publishing.

Conti-Ramsden, G., & Dykins, J. (1991). Mother-child interactions with language-impaired children and their siblings. *Int. J. of Lang. & Com. Disorders*, *26*(3), 337–354.

Ellis Weismer, S., & Hesketh, L. J. (1996). Lexical learning by children with specific language impairment: Effects of linguistic input presented at varying speaking rates. *JSLHR*, *39*(1), 177–190.

Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *JSLHR, 42*, 1249–1260.

Freudenthal, D., Ramscar, M., Leonard, L. B., & Pine, J. M. (2021). Simulating the acquisition of verb inflection in Typically Developing children and children with Developmental Language Disorder in English and Spanish. *Cognitive Science*, *45*(3), 1–29.

Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543.

Goldwater, S., Griffiths, T. L., and Johnson, M. (2011). Producing power-law distributions and damping word frequencies with two-stage language models. *Journal of Machine Learning Research*, 12:2335–2382.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22–44.

Hoeffner, J. H., & McClelland, J. L. (1993). Can a perceptual processing deficit explain the impairment of inflectional morphology in development dysphasia? A computational investigation. In *Proceedings of the 25th Annual Child Language Research Forum* (Vol. 25, pp. 38–49).

Horsborough, K., Moss, T. & Ball, J. (1985). Conversational interaction between mothers and their autistic, dysphasic and normal children. In T. G. Cross & L. M. Riach (Eds.), *Issues and Research in Child Development*, 470–476.

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences*, *2*(7), 240–247.

Johnson, M., Griffiths, T. L., & Goldwater, S. (2007a). Adaptor Grammars: A framework for specifying compositional nonparametric Bayesian models. In *Advances in Neural Information Processing System 19*. 641–648.

Johnson, M., Griffiths, T. L., & Goldwater, S. (2007b). Bayesian inference for PCFGs via Markov chain Monte Carlo. In *Proceedings of NAACL*, Rochester, NY. 139–146.

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *JSLHR*, 53(3), 739–756.

Leonard, L. B. (2014). *Children with specific language impairment*. MIT press.

Leonard, L. B. (1989). Language learnability and specific language impairment in children. *Applied Psycholinguistics*, *10*, 179–202.

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

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, *21*(2), 339–366.

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment, *JSLHR*, *46*, 1138–1153.

McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *JSLHR*, *60*(10), 2891–2905.

O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage*. Cambridge, MA: MIT Press.

Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick Incidental Learning (QUIL) of words by school-age children with and without SLI. *JSLHR*, *38*(2), 434–445.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quart. J. of Exp. Psychol.*, *17*(4), 273–281.

Pitman, J., & Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, *25*(2), 855–900.

Rice, M. L., & Bode, J. V. (1993). GAPS in the verb lexicons of children with specific language impairment. *First Language*, *13*(37), 113–131.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *JSLHR*, *37*(1), 106–122.

Riches, N. G., Tomasello, M. & Conti-Ramsden, G. (2005). Verb learning in children with SLI. *JSLHR*, *48*, 1397–1411.

Tallal, P. & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. *Nature*, *241*, 468–469.

Thomas, M. S. (2005). Characterising Compensation. *Cortex*, *41*(3), 434–442.

Tirosh, E., & Cohen, A. (1998). Language deficit with an attention-deficit disorder: A prevalent comorbidity. *Journal of Child Neurology, 13*, 493–497.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399–433.

Van Kleeck, A., & Carpenter, R. L. (1980). The effects of children's language comprehension level on adults' child-directed talk. *JSLHR*. *23*(3), 546–569.

Weiler, B. (2013). Verb Selection and Past-Tense Morphology, *Topics in Language Disorders*, *33*(2), 152–164.

Windfuhr, K. L., Faragher, B., & Conti-Ramsden, G. (2002). Lexical learning skills in young children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *37*(4), 415–432.