Making Good on BADs

Sarah Payne and Charles Yang
University of Pennsylvania

1 BADs and their Implications

Dressler et al. (2019, henceforth D&al) report several cases of rather unusual patterns in children's morphological acquisition. These patterns are dubbed *Blind-Alley Developments* (BADs) because of their striking deviation from the input as well as their transitory nature, eventually to be replaced with adult-like forms.

D&al discuss two kinds of BADs: we interpret both as misalignments of form and function in children's morphology. Weak BADs have the correct forms but wrong functions. D&al give an example of the German interfix -*e*-. In adult German, for example, -*e*- is used in some noun-noun compounds, but is infrequent and unproductive. However, some children overgeneralize it, forming compounds such as **Luft+e+bon-e* (instead of *Luft+bon-e*, 'air balloons'). Strong BADs, by contrast, get the form wrong altogether. For example, iterativity and ongoingness are expressed by imperfectivtity in Russian. Some children, however, make use of total reduplication as in *njam-njam* ('I'm eating') and *pik-pik* (repeated jumping).

Although uncommon and not affecting every language-learning child, the importance of BADs should be recognized. Rare patterns still require explanation, and they set up boundary conditions on any theory of language and language acquisition. And we suspect that if researchers were to look for BADs with a very fine-toothed comb like D&al, more instances will turn up.

D&al note that BADs pose challenges to both usage-based and universalist theories of language acquisition. We share their skepticism toward usage-based accounts. While children do learn language from the input, the extent to which children's grammar mirrors the input (Tomasello, 2000) has been overstated by usage-based theorists (Yang 2013, Yang 2016, Section 2.1). At the same time, an initial stage of lexically specific grammar is hardly a unique claim of usage-based theories (e.g., Diessel, 2013) but has always been acknowledged by supposedly opposing generative accounts (e.g., Marcus et al., 1992). A further difficulty, though not one unique to usage-based theories, is the absence of a credible learning theory that explains the child's transition from a lexically restrictive stage to an adult-like productive and abstract grammar (Yang, 2015, 2017). Indeed, recent usage-based accounts of morphological acquisition (Engelmann et al., 2019) overproduce patterns with highest token frequency, and thus have similar failings as previous models (e.g., Rumelhart and McClelland, 1986) as minority patterns can be robustly overapplied as well (Elsen, 2002).

We are less certain about D&al's critique of universalist theories. It may be a terminological matter. There are two senses in which the term *universal* can be understood, and they are both part of the theory known as Universal Grammar (Chomsky, 1965, Chapter 1, §5). In the domain of morphology, *formal* universals would characterize general properties of all morphological systems: words have constitutive parts (morphemes), morphemes are composed into structured units, the process of composition may be productive or unproductive, etc. By contrast, *substantive* universals would describe the totality of morphological variations: particular types of morphological units (e.g., roots, infixes, templates, reduplication), the principles by which they combine (e.g., syntax-like locality conditions in Distributed Morphology, Halle and Marantz 1993, the structure and constraints on paradigms, Stump 2001), and their forms and functions (e.g., inflection vs. derivation, the range of semantic relations encoded by morphology). We do not believe that anyone has seriously espoused a substantive theory of universal morphology – or it is wise to do so. This is probably because Bromberger and Halle (1989)'s observation of phonology holds equally for morphology: both are the product of history and other contingencies, and it is difficult to imagine a theory that includes all conceivable morphological patterns in the world – and BADs produced by children. Perhaps D&al are objecting to even the possibility of such a theory; if so, then we agree. At the same time, it seems that almost everyone subscribes to a *formal* theory of universal morphology, e.g., words have internal structures. Admittedly, such a theory is not particularly helpful when it comes to the language-specific particulars: We need another theory – for instance, a learning theory – one which turns the formal into the substance across languages.

D&al's own account of BADs is couched in the Natural Morphology framework (Dressler et al., 1987). The child constructs their morphology making use of a broad range of linguistic and non-linguistic strategies, especially a set of cognitively based universal preferences such as iconicity, transparency, etc. We are sympathetic to this view. Learning morphology clearly requires the intersecting forces of language and cognition: think the role of animacy, shape, size, etc. in noun class(ifier) systems, some of which are highly culturally specific and must be built upon the child's conceptual development. But a constructivist needs to specify the precise mechanism by which the morphological system is built and subsequently refined: How do tendencies and possibilities become actualities? In particular, what exactly leads children astray into BADs – and why not all children? – and how do they manage to escape?

It seems to us, then, that all three theoretical approaches have a central missing piece: a learning-theoretic account of how children learn morphology. We now turn to such a learning-theoretic account, which in our view provides a more complete interpretation of BADs.

2 The Tolerance Principle

Learning a language requires discovering rules that generalize beyond a finite sample of data. The Tolerance Principle (TP) is a theory of how such generalizations are formed. Specifically,

(1) Let a rule R be defined over a set of N items. R is productive if and only if e, the number of items not supporting R, does not exceed θ_N:

$$e \leq \theta_N = \frac{N}{\ln N}$$

If e exceeds θ_N, then the learner will "lexicalize" only these N items and not generalize beyond them: that is, R is unproductive. Yang (2016, p177) introduces a corollary, the Sufficiency Principle, which specifies how generalizations are formed when the "exceptions" are not attested but cannot be regarded as impossible, on the ground that evidence of absence is not absence of evidence. But it is clear that both principles generalize on the number of positive examples: specifically, $N - \theta_N$, a super majority of N. Throughout this paper, we will refer to these variants collectively as the Tolerance Principle (TP).

The TP builds on the intuition that a rule must "earn" productivity by the virtue of being applicable to a sufficiently large number of candidates it is eligible for. For example, if there are 10 examples and all but one (9/10) support a rule, generalization ought to take place. But no one in their right mind would extend a rule on the basis of 2/10: the learner should just memorize the two supporting examples. Productivity is a calibration of regularities and exceptions — crucially with respective to word *types* rather than tokens. For background and motivation for the TP, we refer the reader to (Yang, 2005, 2016) and much subsequent work. Recently, the TP has been implemented as the central component of a computational model of rule learning that proposes, and evaluates, rules over a corpus of data (Belth et al., 2021), in a step toward a discovery procedure envisioned at the beginning of generative grammar (Chomsky, 1957).

By hypothesis, the productivity of a rule is determined by two integer values (N and e) defined over a vocabulary: these are obviously matters of individual vocabulary variation. Thus the TP allows room for variation in the transient stages of language acquisition as well as in the stable grammars of individual speakers. The relationship between N and e, which may change during the course of language acquisition, determines the productivity status of the rule. If e is very low as a proportion of N, then children may rapidly conclude that a rule is productive. Otherwise, a protracted stage of conservatism may ensue, which may be followed by the sudden onset of productivity, as can be seen, famously, in the over-regularization of irregular verbs (Marcus et al., 1992; Yang, 2002). It is also possible that no rule ever reaches the productivity threshold; gaps and other phenomena of ineffability arise (Yang, 2016; Gorman and Yang, 2019). Finally, the stochastic nature of child vocabulary acquisition may result in the productivity of a rule that is unproductive in the adult language, which ultimately corrects itself when a larger and more representative vocabulary is established.

Before seeing how the TP can help account for children's BADs, we correct a misunderstanding on D&al's behalf. They claim "if a BAD, without support in the input, should have miraculously emerged, it should be immediately knocked out by these input-based principles, which is not the case in our example (p134)." This is not so. The calibration of productivity under the TP takes place in the child's *internal* vocabulary, via the values of N and e, rather than those in the external input (e.g., a corpus or a dictionary at the linguist's disposal). Plainly, the child does not, and can not, learn everything they hear in a corpus: their vocabulary actually grows quite slowly. An English learning child, for example, will have a vocabulary of at most 500 words at age two and at most 1,000 words at age three (Fenson et al., 1994; Hart and Risley, 1995). Cross-linguistically, children in even the most favorable language-learning environments only manage a vocabulary of about 1,000 words by three (Hart and Risley, 1995; Szagun et al., 2006; Bornstein et al., 2004), so the threat of instant knockout that D&al are concerned with will not materialize. However, BADs may *eventually* be knocked out once the child's vocabulary reaches a certain size such that the transitory productivity of BADs fails: the TP would thus provide an explanation for their demise.

3 BADs under the Tolerance Principle

To provide a direct TP analysis of D&al's own BAD examples would be difficult: As noted above, doing so requires very accurate vocabulary measures of the children under study. Such individual-level analyses under the TP can and have been done. For instance, in artificial language studies with children (Schuler, 2017; Emond and Shi, 2021), the number of (nonce) words and their morphological forms can be precisely measured after the training stage. The outcome of learning – whether a rule generalizes or not – can be tested with additional nonce words not used in training, akin to a Wug test, and then compared against, and in fact, confirmed, the predictions of the TP based on individual learner's vocabulary counts (N and e). Unfortunately, the vocabulary measures of children in D&al's study are not available: even the full transcripts can only capture a subset of the children's vocabulary. For present purposes, then, we will illustrate the TP approach to BADs with examples from cases for which we do have accessible vocabulary data (or at least their proxies). We nevertheless hope to convey the general method that can be extended to case studies should precise vocabulary measures be available.

3.1 Weak BADs

Weak BADs have always been in the center of attention in language acquisition research albeit under different terminologies such as "analogy." We agree with D&al: despite frequent allusions to analogical forces in language acquisition (e.g., Bybee, 1985; Pinker and Prince, 1988), analogical errors are really BADs as they are quite rare. For instance, while over-regularization errors (e.g., *go-goed*, *hold-holded*) in child English are quite common, analogical errors such as *bite-bote*, *wipe-wope*, and *think-thunk* are vanishingly rare. The most comprehensive empirical study of analogical errors (Xu and Pinker, 1995), in fact, dub these "weird past tense errors" on the basis of their rarity. Xu and Pinker examined over 20,000 past-tense tokens produced by nine children: only forty weird errors (0.02%) were identified, and some of these may well be the result of phonological lenition such as *slep*, which may be the correct form *slept* undergoing the familar process of word-final t/d-deletion. Indeed, a sharp contrast between productive and unproductive processes can be observed in numerous acquisition studies across many languages (Lignos and Yang, 2016), which partly motivated the categorical conception of productivity under the TP.

Xu and Pinker (1995)'s study does reveal a single systematic error of over-irregularization along the line of *sing-sang/sung* in child English, which would correspond to a weak form of BADs. The verb *bring* is over-regularized 32 times in the 4 million word corpus of child English: This is expected. But *brang* appears 6 times and *brung* 5 times. It is difficult to say that all of these are definitively examples of BADs: some dialects of American English do use *brang* and/or *brung* and a few instances can be found in the CHILDES input. Nevertheless, there is little doubt that some children do misgeneralize the pattern of *sing* and *ring* to *bring* to form *brang* and *brung*, as well *swing-swang*.

The *bring-brang* error is a BAD in the weak sense: the child gets the form wrong but the function is correct (i.e., marking past tense). The TP provides a straightforward account for this, as discussed in many case studies of phonological, morphological, and syntactic acquisition (Yang, 2016). Verbs such as *bring*, *sing*, and *ring* are very frequent: in fact, they are among the most frequent 200 verbs in child-directed English. As is well documented (e.g., Goodman et al., 2008), frequency plays an important role in vocabulary acquisition: more frequent words are more likely to be early. Consider the hypothetical case where vocabulary acquisition proceeds in a strictly frequent fashion. By the time that the child learns 200 verbs, they will have learned 76 irregular verbs (based on CHILDES): English irregular verbs are notoriously frequent. Note that 76 irregulars are far too many to support the productivity of -*ed* rule despite its coverage of a majority of 127 verbs. According to the TP, a rule can generalize over a vocabulary of 200 items only if the number of its exceptions falls below $200/\ln 200 = 37$. However, the learner, as in the computational model of Belth et al. (2021), can find a productive rule defined over a subset of the verbs when the entire vocabulary fails to yield a productive rule. Specifically, the rule ɪ → æ / __ ŋ reaches productivity because it works for two (*sing* and *ring*), and the one exception of *bring* can be tolerated. In fact, the rule can enjoy an extended period of productivity. Among the top 500 verbs, the two exceptions (*bring* and *swing*) cause no difficulty for the remaining three. A case of BAD.

But the productivity of ɪ → æ / __ ŋ won't last too long. Consider the relevant words when the vocabulary reaches 800:

(2) bring, *fling* , **ring**, **sing**, **spring**, *sting*, *swing*, wing

There are now 8 verbs that end in /ɪŋ/ but their past tense forms are scattered all over the place: the three in boldface change vowel to /æ/, the three in italic change the vowel to /ʌ/, one is completely idiosyncratic (*bring*), and the last one is regular (*wing*). Everyone loses because no one is numerically dominant enough to tolerate the rest. At this point, children will stop the generalization and memorize the past tense of irregular "ɪŋ" verbs by rote. The absence of productivity for this class of verbs accounts for the overall rarity of "weird past tense" errors: D&al's BADs. Importantly, adults also treat this class as unproductive and do not generalize the irregular patterns. The past few decades have seen verbs such as *bing* (Microsoft search engine) and *bling* enter into the English vocabulary: they are uniformly inflected with the regular "-ed".

Of course no child learns words on a strictly frequency-determinant fashion. The actual vocabulary is surely a stochastic sample of input words: the fortuitous condition for the rise of *bring-brang* is only met for some children. But there is nothing unprincipled about our account: every child has a vocabulary, and the only limiting factor is our ability to measure it precisely. Once we have N's and e's, TP predictions are completely mechanical and unambiguous. BADs therefore receive an explanation from start to finish: It is simply the rise and fall of productivity stemmed in the child's changing vocabulary. The same approach holds for language change: see Ringe and Yang (2022) for how the TP accounts for the historical productivity of the *ing* class and how it was lost over time.

3.2 Strong BADs

A few qualifying words on strong BADs first. In our understanding, the strong cases of BADs in D&al's study are not children's inventions out of thin air but are in fact grounded somewhere in the input data. Consider the case of reduplication in child Russian and Greek. While it is true that reduplication is not used to express, respectively, iterativity/ongoingness and the subjunctive in these languages, reduplication, as a formal pattern, does appear in both languages. It is used in Russian to express, among others, intensification (Israeli, 1997), and is used in Modern Greek to express a complex range of semantic and pragmatic information (Kallergi, 2015). While children's BADs are clearly different from these, they are appropriately described as misalignment between an extant morphological form and its functions.

Viewed under these lights, then, strong BADs may not be as rare or atypical as D&al suggest. Quite a lot of research in language acquisition, including morphological acquisition, has identified such misalignments of form and function as an especially revealing type of errors. One prominent example is the extensively studied phenomenon of Root Infinitives (RI) whereby children frequently use non-finite verbal forms in matrix sentences that require a finite form (Rizzi, 1993; Hoekstra and Hyams, 1998; Wexler, 1998; Legate and Yang, 2007). In French, for instance, non-finite matrix clauses are restricted to particular discourse situations that have a time reference as shown in the following dialogue (3) and questions (4), drawn from Rasetti (2003):

(3) a. Que veux-tu faire?
What want you do$_{\text{INF}}$
'What do you want to do?'

b. Partir
leave$_{\text{INF}}$
'(I want) to leave.'

(4) a. Comment lui dire cela?
How him$_{\text{DAT}}$ say$_{\text{INF}}$ this
'How to say this to him?'

b. Moi faire ceci? Jamais!
Me do$_{\text{INF}}$ that? Never!

French-learning children, however, use non-finite matrix verbs in broader contexts and not infrequently – and crucially, in *declarative* sentences, over 10% of the time for the three children in Rasetti's study. Children's use of RIs falls into two broad classes in French: it may have a modal interpretation (5a) or as description (5b).

(5) a. manger # maman [= reaching for some chocolate]
eat$_{\text{INF}}$ mommy.

b. cacher le caryon # voilà!
hide$_{\text{INF}}$ the pencil, there!

The RI phenomena, then, appears to be a strong case of BAD: the child erroneously maps an input form (non-finite verbs) to a non-target function (matrix usage in stand-alone declarative sentences). But unlike the examples of BADs in D&al's paper, which struck only select children, RI has been observed in all children learning tense-marking languages with quantitative cross-linguistic differences regarding the extent and duration of non-target form use.

RIs, too, can be accounted for under the TP. Before learning the correct mappings from form to function, children must first learn which functions are realized – in this case, which morphological features are marked – in their language. Chinese, for example, does not mark tense, while English does. English, on the other hand, does not contrast first and second person, while Spanish and French do. Which subsets of morphological features are marked, and are thus eligible to be mapped to form, vary across languages and must be learned from the input. Some proposals regarding the RI phenomenon have suggested that it results from under-specified features in the Tense node (e.g. Hoekstra and Hyams 1998). Our proposal, which draws on Payne (2022), is in a similar spirit: we suggest that the RI emerges before the child has seen sufficient evidence for tense marking; at this stage, tense is essentially under-specified, and children may supply verbal forms inappropriate for the function, i.e., non-finite forms for finite context.

The TP can be straightforwardly applied to the problem of learning which features are marked in a language. Much work on morphological acquisition (e.g. Kim 2015; Kim and Sundara 2021; Marquis and Shi 2012) has demonstrated that children's ability to segment inflectional morphology and relate inflected forms to their stems emerges early in acquisition – as early as 8 months in English. At the same time, the Principle of Contrast (Clark and MacWhinney, 1987) provides a (perhaps innate) hypothesis that different forms will correspond to different meanings, or, in the language of the present discussion, different functions. We suggest that the child may combine their early ability to segment and relate inflected forms with the Principle of Contrast in order to detect and make use of *collisions* – a single stem appearing in multiple inflected forms (e.g. *walk-walked*). Such collisions may then be used to determine which features are marked – and thus which functions may be mapped to form – in their language (e.g. *walk-walked* provides evidence for ± PAST).

Yet a single collision is likely not sufficient evidence that a language realizes the corresponding features: English, for example, distinguishes first and second person in *I am - you are* but this distinction does not generalize (c.f. *I walk - you walk*). Here, we may straightforwardly apply the TP: once the child learns which function is realized by individual collisions (e.g. ± PAST above), they may employ the TP to determine whether there is sufficient evidence for the realization of that function across the language. More formally, if we encounter a collision between inflected form A and inflected form B, where A is less frequent, then we determine the morphological features that are realized by this collision. We then search the lexicon for other words that have appeared in inflected form A. If enough of these also have a collision that realizes the same features (where enough is defined by the TP), then we learn that these features are marked in the language. We focus on the less frequent of the two inflected forms because the number of collisions between A and B will be higher relative to the total count of A than the total count of B, giving it the best chance of passing the TP. We refer the reader to Payne (2022) for the details and computational implementation of this approach.

Learning under this proposal is recursive: the child may learn, for example, that ± PARTICIPLE is marked in English, and then that ± 3, SG is marked for − PARTICIPLE forms, and so on. The order in which feature marking is learned, which is dependent on the vocabulary the child has acquired incrementally, can be translated into a feature hierarchy representation familiar in morphological theories. For languages where agreement emerges before tense, this necessarily means that tense will emerge separately for each agreement. While such an approach may at first seem counter-intuitive, it may actually provide a direct explanation for cross-linguistic differences in RI phenomena. Several accounts of RI (e.g. Phillips 1995; Guasti 2002) have suggested that the length and frequency of the RI stage is inversely correlated with the morphological "richness" of the language: languages with richer inflectional morphology yield shorter RI stages. While the idea of "richness" itself is not a precise or sufficient explanation, Legate and Yang (2007) build on these intuitions by suggesting that it is not the "richness" of the paradigm that matters, but rather the relative amount of forms the child encounters that reward a tense-marking grammar. In the case of Spanish, English, and French, they show that the child will encounter tense-rewarding forms most in Spanish and least in English, which matches with the relative length of the RI period (longest in English and shortest in Spanish).

Our proposal offers an alternate explanation: In English, French, and Spanish, agreement generally emerges before tense (Brown, 1973; Berko, 1958; Prévost, 2009; Montrul, 2004), which means that our model will subdivide the input based on agreement and learn tense marking within each agreement node separately. At the same time, the TP tolerates relatively more exceptions for smaller values of N: when $N = 10$, $\theta_N = 4.3$, or about 43% of N. By contrast, when $N = 100$, $\theta_N = 21.7$, or about 22% of N. This means that it will be easier to learn tense marking over smaller N, or equivalently, that tense will emerge more quickly in languages that subdivide more based on agreement. This is indeed the case: English marks only one agreement (± 3,SG), while Spanish marks 5 or 6, depending on the dialect. French falls between, with the first and second plural being the only agreements that are realized phonologically (Legate and Yang, 2007). Our proposal straightforwardly predicts that tense-marking will emerge most quickly in Spanish, followed by French and then English, based solely on properties of subdivision under the TP. These predictions match with developmental findings regarding both age of acquisition and length of the RI stage (Brown, 1973; Phillips, 1995; Montrul, 2004; Legate and Yang, 2007; Prévost, 2009).

It is important to note that learning *what* functions are realized – or equivalently, what features are marked – in a language does not mean that the child has yet learned *how* to map these functions to form. At the stage when the child only knows agreement marking, it is unlikely that any morphological process will be productive since the processes differ based on tense. At this point, for example, the Spanish-learning child would group *amas, amabas, amaras, amaste*, etc. because all forms mark the second singular, but no single morphological process can derive all four forms. This explains why the child may produce the RI, as well as its gradient nature: if there is not yet a productive rule to inflect a given agreement form, the child must rely entirely on lexicalized inflected forms with the desired agreement. In the absence of such forms, and the absence of a productive rule for inflection, it is reasonable to expect that the child may resort to an RI form as long as it marks agreement correctly. Once tense and other markings do emerge, the child will then be able to map these functions to form, and weak BADS may emerge. The TP can thus provide a direct account for the emergence, gradience, and subsequent demise of the RI, as well as cross-linguistic differences in this strong BAD.

4 Does Everything Go?

We offer the TP as a *formal* universal of language (including morphology). It is possible that the TP is rooted in the non-linguistic component of cognition: a general principle of learning and generalization although empirical work in other domains will be necessary. Any substantive morphological rule, as long as it is clearly defined, can receive a TP verdict via a calculation with two cardinality values. Children may postulate many rules during the course of acquisition: many survive in the long run so we have a relatively stable grammar across individuals and BADs are just premature ones that do not.

But where does the substance of morphology come from? Presumably this is D&al's Feyerabendian question "Does everything go?" Natural Morphology likely constitutes part of the answer. The fact that children, unprompted, use reduplication to express iterativity seems to be grounded in cognitive factors such as iconicity: repetition is repetition. But we doubt a complete theory of substance is ever forthcoming. The usual suspects – sounds and meanings – no doubt participate in the construction of morphology, but D&al make it clear that all sorts of individualistic factors come into play as well. Children learn different words, they latch onto different aspects of the world around them, and every little perturbation, such as a morpheme segmentation error (Peters, 1983), can give rise to BADs.

There will never be a theory of everything as that would amount to the totality of human imagination meshed together all the vagaries of life and experience. We cannot predict what a child comes up on their way to the final grammar; we can, however, give a definite answer whether it actually goes.

References

Belth, C., Payne, S., Beser, D., Kodner, J., and Yang, C. (2021). The greedy and recursive search for morphological productivity. In *Proceedings of CogSci 2021*.

Berko, J. (1958). The child's learning of english morphology. *Word*, 14(2-3):150–177.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Pêcheux, M.-G., Ruel, J., Venuti, P., and Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, dutch, french, hebrew, italian, korean, and american english. *Child development*, 75(4):1115–1139.

Bromberger, S. and Halle, M. (1989). Why phonology is different. *Linguistic Inquiry*, 20(1):51–70.

Brown, R. (1973). *A first language.* Harvard University Press.

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

Chomsky, N. (1957). *Syntactic structures.* Mouton, The Hague.

Chomsky, N. (1965). *Aspects of the theory of syntax.* MIT Press, Cambridge, MA.

Clark, E. V. and MacWhinney, B. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition*, pages 1–33.

Diessel, H. (2013). Construction grammar and first language acquisition. In Hoffmann, T. and Trousdale, G., editors, *The Oxford handbook of construction grammar*, pages 347–364. Oxford University Press, Oxford.

Dressler, W. U., Christofidou, A., Gagarina, N., Korecky-Kröll, K., and Kilani-Schochd, M. (2019). Morphological blind-alley developments as a theoretical challenge to both usage-based and nativist acquisition models. *Italian journal of linguistics*, 31(2):107–140.

Dressler, W. U., Mayerthaler, W., Panagl, O., and Wurzel, W. U. (1987). *Leitmotifs in natural morphology*. Benjamins Amsterdam.

Elsen, H. (2002). The acquisition of German plurals. In *Morphology 2000: Selected Papers from the 9th Morphology Meeting, Vienna, 25-27 February 2000*, volume 218, page 117. John Benjamins Publishing.

Emond, E. and Shi, R. (2021). Infants' rule generalization is governed by the Tolerance Principle. In Dionne, D. and Vidal Covas, L.-A., editors, *Proceedings of the 45nd annual Boston University Conference on Language Development*, pages 191–204.

Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., Theakston, A., and Lieven, E. (2019). How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology*, 110:30–69.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., and Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, pages i–185.

Goodman, J. C., Dale, P. S., and Li, P. (2008). Does frequency count? parental input and the acquisition of vocabulary. *Journal of child language*, 35(3):515–531.

Gorman, K. and Yang, C. (2019). When nobody wins. In Rainer, F., Gardani, F., Luschützky, H. C., and Dressler, W. U., editors, *Competition in inflection and word formation*, pages 169–193. Springer, Berlin.

Guasti, M. T. (2002). *Language acquisition: The growth of grammar*. MIT press.

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. and Keyser, S. J., editors, *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, pages 111–176. MIT Press, Cambridge, MA.

Hart, B. and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing, Baltimore, MD.

Hoekstra, T. and Hyams, N. (1998). Aspects of root infinitives. *Lingua*, 106(?):81–112.

Israeli, A. (1997). Syntactic reduplication in russian: A cooperative principle device in dialogues. *Journal of Pragmatics*, 27(5):587–609.

Kallergi, H. (2015). *Reduplication at the Word Level*. De Gruyter Mouton.

Kim, Y. J. (2015). *6-month-olds' segmentation and representation of morphologically complex words*. University of California, Los Angeles.

Kim, Y. J. and Sundara, M. (2021). 6–month–olds are sensitive to english morphology. *Developmental science*, 24(4):e13089.

Legate, J. A. and Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, 14(3):315–344.

Lignos, C. and Yang, C. (2016). Morphology and language acquisition. In Hippisley, Andrew R. abd Stump, G., editor, *The Cambridge handbook of Morphology*, chapter 28, pages 765–791. Cambridge University Press, Cambridge.

Marcus, G., Pinker, S., Ullman, M. T., Hollander, M., Rosen, J., and Xu, F. (1992). *Overregularization in language acquisition*. Monographs of the Society for Research in Child Development. University of Chicago Press, Chicago.

Marquis, A. and Shi, R. (2012). Initial morphological learning in preverbal infants. *Cognition*, 122(1):61–66.

Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*, volume 37. John Benjamins Publishing.

Payne, S. (2022). When collisions are a good thing: the acquisition of morphological marking. Bachelor's thesis, University of Pennsylvania.

Peters, A. M. (1983). *The units of language acquisition*. Cambridge University Press, Cambridge.

Phillips, C. (1995). Syntax at age two: Cross-linguistic differences. *Language Acquisition*, 17(1-2):70–120.

Pinker, S. and Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1):73–193.

Prévost, P. (2009). *The acquisition of French: The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*, volume 51. John Benjamins Publishing.

Rasetti, L. (2003). *Optional categories in early French syntax: A developmental study of root infinitives and null arguments*. PhD thesis, University of Geneva.

Ringe, D. and Yang, C. (2022). The threshold of productivity and the irregularization of verbs in Early Modern English. In Los, B., Honeybone, P., and Trousdale, G., editors, *English historical linguistics: Change in structure and meaning*, pages 91–111. John Benjamins, Amsterdam.

Rizzi, L. (1993). Some notes on linguistic theory and language development: The case of root infinitives. *Language acquisition*, 3(4):371–393.

Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tenses of English verbs. In McClelland, J. L., Rumelhart, D. E., and the PDP Research Group, editors, *Parallel distributed processing: Explorations into the microstructure of cognition. Volume 2: Psychological and biological models*, pages 216–271. MIT Press, Cambridge, MA.

Schuler, K. (2017). *The acquisition of productive rules in child and adult language learners*. PhD thesis, Georgetown University, Washington, D.C.

Stump, G. (2001). *Inflectional morphology: A theory of paradigm structure*. Cambridge University Press, Cambridge.

Szagun, G., Steinbrink, C., Franik, M., and Stumper, B. (2006). Development of vocabulary and grammar in young german-speaking children assessed with a german language development inventory. *First Language*, 26(3):259–280.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3):209–253.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, 106(1):23–79.

Xu, F. and Pinker, S. (1995). Weird past tense forms. *Journal of Child Language*, 22(3):531–556.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford University Press, Oxford.

Yang, C. (2005). On productivity. *Linguistic Variation Yearbook*, 5(1):333–370.

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 110(16):6324–6327.

Yang, C. (2015). Negative knowledge from positive evidence. *Language*, 91(4):938–953.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language*. MIT Press, Cambridge, MA.

Yang, C. (2017). Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition*, 24(2):100–125.