# Chapter 11

# How to wake up irregular (and speechless)

Charles Yang
University of Pennsylvania

I suggest that morphological defectiveness arises when the learner fails to discover a productive/default process in a morphological category. The detection of productivity, or lack thereof, can be accomplished by the Tolerance Principle, a simple mathematical model of language learning and generalization. In this paper, I show that the absence of **amn't*, the negative contracted form of *am*, in most English dialects can be predicted on purely numerical basis. Implications for language acquisition, variation, and change are also discussed.

## 1 From Irregular Verbs to Productivity

In my first linguistics talk, which was also my job interview at Yale, I proposed that English irregular past tense is not learned by forming associations between the stem and the inflected form, contrary to the dominant view in the psychological study of language (Rumelhart & McClelland 1986; Pinker 1999). Rather, irregular past tense is generated by *morpholexical* rules. These rules do not generalize beyond a fixed list but are rules nevertheless, in the sense that they take the stem (e.g., *think*) as the input and generate an output (e.g., *thought*), the inflection, via a computational process of structural change (e.g., "Rime → /ɔt/"). I was approaching the problem as a computer scientist: rules are most naturally realized as a list of if-then statements, for regulars and irregulars alike, which turns out to be the approach taken throughout the history of linguistics (Bloch 1947; Chomsky & Halle 1968; Halle & Marantz 1993) including Steve's own work (1973; 1992). There is in fact developmental evidence for the rule-based approach when I reanalyzed the past tense acquisition data purportedly confirming the associationist account (Yang 2002b).

I supposed Steve was at least somewhat persuaded by the argument; a few months later I got the job. But he did wonder aloud after the talk, with a quizzical frown-cum-smile that only he can manage: "But how does a rule wake up in the morning and decide to be irregular?"

Charles Yang. 2017. How to wake up irregular (and speechless). In Claire Bowern, Laurence Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond)*, 211–233. Berlin: Language Science Press. DOI:10.5281/zenodo.495447

Indeed. Since words do not wear tags of (ir)regularity, any morphological theory that recognizes regularity and irregularity, which is pretty much everything on the market, must say something about how a rule or process wakes up to be irregular. In fact, theories that reject such a categorical distinction (e.g., Hay & Baayen 2003; McClelland & Patterson 2002) ought to be off market. Children's morphological productivity is strongly discrete; see Yang 2016: Chapter 2 for a cross-linguistic review. Their errors are almost exclusively over-regularizations of productive rules. This is quite well known thanks to the past tense debate: for example, the past tense of *hold* sometimes surfaces as *holded*, with the "-d" rule (Marcus et al. 1992). What is not widely known and even less appreciated is the near total absence of over-*irregularization* errors, despite frequent anecdotes to the contrary (e.g., *bite-bote*, *wipe-wope*, *think-thunk*, etc.; Bowerman 1982; Bybee 1985; Pinker 1999). These errors are sufficiently rare, occurring in about 0.2% of English-learning children's past tense use, that Xu & Pinker (1995) dub them "weird past tense errors". Not a single instance of *bote*, *wope*, *thunk*, or many conceivable analogical patterns can be found in the millions of child English words in the public domain (MacWhinney 2000). The distinction between regular and irregular rules was in fact observed in Berko's (1958) original Wug test. While children were quite happy to add "-d" to novel verbs such as *rick* and *spow*, only one out of eighty six subjects irregularized *bing* and *gling*, although adults are often prone to form irregular analogies in an experimental setting.[1]

So Steve's question sent me on a long quest. To maintain that both regulars and irregulars are computed by rules, I needed a story of how children separate out productive and unproductive rules so precisely and effortlessly. Although a solution was worked out shortly after (Yang 2002a), it took me many years to fully recognize the scope of the productivity problem – one of the "central mysteries" in morphology (Aronoff 1976: 35) – and the challenges it poses.

At a first glance, it doesn't seem difficult to give an answer for English past tense. The rule "add -d" covers most verb types in the language and can thus be deemed regular, as "statistical predominance" has always been regarded as the hallmark of the default (e.g., Nida 1949: 14). But this is surely too simplistic when crosslinguistic and psychological factors are taken into account. More concretely, at least four empirical problems, each of which is illustrated with a familiar example in (1), fall under the productivity problem.

(1) a. English past tense: That a default rule is learned abruptly and results in over-regularization, after a protracted stage of rote memorization (Marcus et al. 1992; Yang 2002b).

b. English stress: That the grammar of English stress (Chomsky & Halle 1968; Hayes 1982; Halle & Vergnaud 1987) is not trochaic with a list of lexical exceptions despite an overwhelming majority of English words bearing stress on the first syllable (Cutler & Carter 1987; Legate & Yang 2013).

[1] This suggests that the Wug test and similar methods such as rating have task-specific complications and should not be taken as a direct reflection of an individual's morphological knowledge; see Schütze 2005 and Yang 2016 for discussion.

c. German noun plurals: That a suffix ("-s") can be the productive default despite coverage of fewer nouns than any of its four competitors (Clahsen et al. 1992; Wiese 1996).
d. Russian gaps: That morphological categories needn't and sometimes do not have a default, as illustrated by the missing inflections of certain Russian verbs in the 1st person singular non-past (Halle 1973).

In Yang (2016), I propose a model of productivity, the Tolerance Principle, which provides a unified solution for the problems in (1), as well as similar problems that involve inductive learning in phonology, syntax, and language change. In this paper, I revisit Steve's question which, in a significant way, drove this project forward. My focus is on a topic that has featured prominently in Steve's recent research: morphological gaps and the nature of defectiveness in word formation (e.g., Anderson 2008; 2010b).

## 2 The Tolerance Principle

The development of the Tolerance Principle started as a purely formal conjecture: How would one represent a rule (*R*) and the exceptions of that rule (e.g., a set of words $w_1$, $w_2$, ..., $w_n$)? If one is committed to a mechanistic account of the matter – like a computer programmer, for instance – perhaps the only way to encode rules and exceptions is through a set of conditional statements:

(2) If $w = w_1$ Then ...
If $w = w_2$ Then ...
...
If $w = w_e$ Then ...
Apply *R*

This of course immediately recalls the Elsewhere Condition, ever present in linguistics since Pāṇini (Anderson 1969; Aronoff 1976; Kiparsky 1973; Halle & Marantz 1993). In particular, the data structure in (2) entails that in order for a (productive) rule to apply to a word, the system must scan through a list to ensure that it is *not* one of the exceptions ($w_1$, $w_2$, ..., $w_e$).

There is something perverse about (2). For example, to produce *walked*, one must scan through the irregular verbs to make sure that *walk* is not found on the list. But a moment of reflection suggests that the Elsewhere Condition makes perfect sense. The alternative to listing the irregulars would have to be listing the regulars. One can imagine assigning each regular verb a flag, which immediately triggers the application of the "add -d" rule. But that would imply that the morphological status of *every* word must be committed to special memory; the irregulars as well, since they are by definition unpredictable. Perhaps even more surprisingly, there is broad behavioral support for the irregulars-first-regulars-later representation of rules; see Yang 2016: Chapter 3 for review. The psycholinguistic evidence comes from real-time processing of words and morphology. When irregulars and regulars are suitably matched for various factors (e.g., stem and surface frequency) that affect the speed of processing, irregulars are recognized and produced significantly *faster* than regulars – which is consistent with the algorithmic interpretation of the Elsewhere Condition as a computational model of language processing.

From (2), then, we can develop an empirically motivated cost-benefit calculus for the price of exceptions. Specifically, words that fall under a productive rule must "wait" for the exceptions to be processed first: the more exceptions there are, the longer the rule will have to wait. Under very general assumptions about word frequencies, we can prove:

(3) *Tolerance Principle*
Suppose a rule $R$ is applicable to $N$ items in a learner's vocabulary, of which $e$ are exceptions that do not follow $R$. The sufficient and necessary condition for the productivity of $R$ is:

$$e \leq \theta_N \text{ where } \theta_N := \frac{N}{\ln N}$$

The Tolerance Principle requires two input values, $N$ and $e$, and returns the productivity status of a rule. Its application requires a well-defined rule such that $N$ and $e$ can be measured, by the child learner during language acquisition and by the researcher when studying linguistic productivity. To learn the structural description of a rule, typically in the form of "X $\longrightarrow$ Y", one will need to invoke inductive learning models such as those studied in artificial intelligence, cognitive science, and indeed linguistics (e.g., Chomsky 1955). Almost all inductive models form generalizations over specific learning instances and try to discover the shared characteristics of individual elements associated with a shared pattern. For example, suppose two good baseball hitters can be described with feature bundles [+red cap, +black shirt, +long socks] and [+red cap, +black shirt, +short socks]. The rule "[+red cap, +black shirt] $\longrightarrow$ good hitter" will follow, as the shared features (cap, shirt) are retained and the conflicting feature (sock) is neutralized. Obviously, the application of inductive learning must encode the structural constraints on the human language faculty and other cognitive systems implicated in language acquisition (Chomsky 1965). While it is clear that the properties of human language are far from arbitrary, it remains an open question to what extent they reflect a unique system of Universal Grammar (e.g., Merge; Berwick & Chomsky 2016) or general principles of cognition and learning that show continuities with other domains and species; see Yang 2004; Chomsky 2005; Yang et al. 2017 for general discussions.

Table 1 provides some sample values of $N$ and the associate threshold value $\theta_N$.

The apparently, and perhaps surprisingly, low threshold has interesting implications for language acquisition. Most importantly, it suggests that all things being equal, smaller vocabulary (smaller values of $N$) can tolerate relatively more exceptions. That is, productive rules are *more* detectable when learners have *less* experience with a language, especially when they have a small lexicon that only consists of relatively high frequency words. This may explain children's remarkably early command of the main ingredients

Table 1: The tolerance threshold for rules of varying sizes

| $N$ | $\theta_N$ | % |
|---|---|---|
| 10 | 4 | 40.0 |
| 20 | 7 | 35.0 |
| 50 | 13 | 26.0 |
| 100 | 22 | 22.0 |
| 200 | 38 | 19.0 |
| 500 | 80 | 16.0 |
| 1,000 | 145 | 14.5 |
| 5,000 | 587 | 11.7 |

of language (Yang 2013), as well as the reason why maturational constraints may aid rather than hamper language acquisition (Newport 1990); see Yang 2016: Chapter 7 for extensive discussion.

The Tolerance Principle has proved highly effective. In Yang (2016), it was applied almost 100 times, making accurate productivity predictions across many languages and domains using only corpus statistics. Furthermore, experimental studies in collaboration with Kathryn Schuler and Elissa Newport have found near categorical confirmation for the Tolerance Principle in artificial language studies with young children (Schuler, Yang & Newport 2016). Some of these robust results are unexpected. This is because the derivation in (3) makes use of numerical approximations that only hold when $N$ is large. In the empirical case studies, however, the value of $N$ is often very modest (e.g., 8 or 9 in the artificial language studies) as it refers to the number distinct lexical items in a morphological category. For the moment, I put these questions aside and return to the problems in (1): the low threshold of exceptions provides just the right approach to the productivity problem across languages.

Consider first the acquisition of English past tense. Through an inductive process illustrated earlier, the phonological diversity of the regulars will quickly establish that any verb can take the "-d" suffix. Its productivity will be determined by the total number of verbs ($N$) and the irregulars ($e$) in the learner's vocabulary. The same consideration applies to the irregular rules. For instance, the seven irregular verbs *bring*, *buy*, *catch*, *fight*, *seek*, *teach*, and *think* all follow the stem change "ought". Such a mixed bag of phonological shapes will also yield an all-inclusive rule, as shown by computational implementations (Yip & Sussman 1998). But the "ought" rule will fare terribly. It only works for seven items, with hundreds and thousands of exceptions, far exceeding the tolerance threshold. As a result, the rule will be relegated to lexicalization. Other irregular patterns can be analyzed similarly: as I show elsewhere (Yang 2016: Chapter 4), they all wake up nonproductive in the morning, thereby accounting for the near total absence of over-irregularization errors (Xu & Pinker 1995).

Following the same logic, we can see that the emergence of the "-d" rule will require a long period of gestation. Although children can quickly induce its structural description – using no more than a few dozen verbs (again Yip & Sussman 1998) – their early verbs will contain many irregulars. Of the top 200 verbs inflected in the past tense (MacWhinney 2000), 76 are irregulars. Because $\theta_{200}$ is only 37, it follows that children who know some 200 most frequent verbs cannot recognize the productivity of "-d" despite its "statistical predominance". During this period of time, even though verbs may be produced with the "-d" suffix, they are in effect irregular: the suffix has no productivity and does not extend beyond a fixed list rote-learned from the input. The telltale evidence for productivity comes from the first attested overregularization errors (Marcus et al. 1992). For individual learners with reasonably complete records of language development, the Tolerance Principle can help us understand why the regular rule becomes productive at that exact moment it did. For example, "Adam", the poster child of English past tense research (Pinker 1999), produced his first over-regularization error at 2;11: "What dat feeled like?" In the transcript of almost a year prior to that point, not a single irregular verb past tense was used incorrectly. It must be that by 2;11, Adam had acquired a sufficiently large number of regulars to overwhelm the irregulars. To test this prediction, I extracted every verb stem in Adam's speech until 2;11. There are $N = 300$ verbs in all, out of which $e = 57$ are irregulars. This is very close to the predicted $\theta_{300} = 53$, and the small discrepancy may be due to the under-sampling of the regulars, which tend to be less frequent and thus more likely missing from the corpus. The critical point to note here is that Adam apparently needed a filibuster-proof majority of regular verbs to acquire the "-d" rule: this is strongly consistent with the predictions of the Tolerance Principle as illustrated in Table 1.

The problems of English stress and German plurals in (1) are similar. In the English case, the assignment of stress to the first syllable may be transiently productive when the child has a very small vocabulary (Kehoe & Stoel-Gammon 1997; Legate & Yang 2013). But it will fail to clear the tolerance threshold when the vocabulary reaches a modest size: even 85% of coverage is not sufficient for larger values of $N$ (e.g., 5000; Table 1). In the German case, none of the five plural suffixes can tolerate the other four as exceptions, not least the "-s" suffix, which covers the smallest set. In both cases, the learner will carry out *recursive* applications of the Tolerance Principle. When no rule emerges as productive over the totality of a lexical set, the learner will subdivide it along some linguistic dimension, presumably making use of constraints on language and other cognitive systems, and attempt to discover productive rules within. Such a move, while more complex, is always more likely to yield productive rules: again, smaller $N$'s that result from subdividing the lexicon tolerate a relatively higher proportion of exceptions than larger $N$'s. For the acquisition of stress, dividing words into nouns and verbs and taking the syllabic weight into account, as prescribed by all modern metrical theories, lead to productive rules of stress assignment, an outcome that accords well with both structural and behavioral findings (Ladefoged & Fromkin 1968; Baker & Smith 1976; Kelly 1992; Guion et al. 2003). The study by Legate & Yang (2013) also reveals important differences between theories of stress in their statistical coverage of the English lexicon: while all theories handle a great majority of English words, only the theory of Halle 1998 clears the tolerance threshold of exceptions. For the acquisition of German plurals, the move is to subdivide the nouns by grammatical gender as well as phonological conditions, similar to certain theoretical approaches to German morphology (e.g., Wiese 1996). The *-s* suffix indeed survives as the default because the other suffixes are productive with more restrictive domains of nouns.

The emergence of morphological gaps is a logical outcome of the Tolerance Principle, which constitutes the topic of the present study. When a rule wakes up irregular, the learner must learn, from positive evidence, the inflected form for each word. Failing to hear a particular inflected form will render the speaker speechless when that form is needed.

# 3 Why Am+Not ≠ Amn't?

## 3.1 Conditions on Gaps

Many current theories of morphology, including Distributed Morphology (for which see Halle & Marantz 1993), Optimality Theory (Prince & Smolensky 2004), Dual-Route Morphology (Pinker 1999; Clahsen 1999), Network Morphology (Brown & Hippisley 2012), Paradigm Function Morphology (Stump 2001) and others, invoke the notion of competition, which by design results in a default or winning form (at least in the inflectional domain). This architectural feature of the theories is inherently incompatible with the existence of morphological gaps, which are quite widespread across languages (Baerman, Corbett & Brown 2010). The Tolerance based approach, while also competition based (through the Elsewhere Condition), does not stipulate a default or productive rule as a primitive in the theoretical machinery. Rather, the presence or absence of a productive rule is the outcome of language acquisition, to be determined by children through the composition of the linguistic data. More specifically, the Tolerance Principle provides the following corollary (Yang 2016: 142):

(4) *Conditions on gaps*
Consider a morphological category $C$ with $S$ alternations, each affecting $N_i$ lexical items ($1 \leq i \leq S$), and $\sum_i N_i = N$. Gaps arise in $C$ only if:

$$\forall i, 1 \leq i \leq S, \sum_{j \neq i} N_j > \theta_N$$

That is, none of the alternations ($S_i$) in $N$ are sufficiently numerous to tolerate all the rest ($\sum_{j \neq i} N_j$) as exceptions: no productive alternation will be identified. The speaker must hear the morphological realization of every word in $C$; if any is to slip through the cracks, a defective gap appears. I should note that in the conception and application of the Tolerance Principle, the terms such as "category" and "alternation" are meant to be general and not restricted to morphology per se. For instance, "category" can be interpreted as any well-defined structural class with a finite number of elements (phonemes, words, morphosyntactic structures, the directionality of a finite number of functional heads, etc.), and "alternation" can be understood as any outcome of a computational process defined over such a class. The Tolerance Principle asserts that in order for a productive pattern to emerge, one of the alternations must be statistically dominant. Elsewhere I have studied several well-known gaps in English, Polish, Spanish, and Russian (Yang 2016: Chapter 5). Their presence is predictable entirely on numerical ground, requiring nothing more than tallying up the counts of the lexical items subject to each alternation. In what follows, I provide a Tolerance Principle account of another much-studied instance of a defective paradigm.

## 3.2 The Statistics of N't Gaps

In many dialects of English, *n't* is not permitted to contract onto auxiliary verbs such as *am* and *may*, as seen in the unavailability of, for example, "*I amn't tired" and "*You mayn't do that" (e.g., Anderwald 2003a; Bresnan 2001; Broadbent 2009; Frampton 2001; Hudson 2000; Zwicky & Pullum 1983). Following Zwicky & Pullum (1983), I will assume that the contracted negative *n't* is an inflectional affix. The question is why *n't* cannot attach to all auxiliary verbs residing in the Tense node. From the perspective of the Tolerance Principle, the emergence of gaps must result from a critical mass of exceptions to the contraction process.

Let us consider the behavior of the auxiliary hosts for *n't*. Zwicky & Pullum (1983: p508) provide a near comprehensive list, which I revise with some additional information in Table 2.

Table 2 provides the frequencies of the auxiliary verbs and their negation in both uncontracted and contracted forms in the 520-million-word Corpus of Contemporary American English (COCA; Davies 2008). Given the heterogeneity of the textual sources, a handful tokens of *amn't* and *mayn't* can be found albeit at very low frequencies. The *n't*-contracted forms of *shall* and *dare* – *shan't* and *daren't* – are also impossible for most American English speakers but are included here for completeness. Although *shan't* is often perceived as a stereotypically British English feature, it seems to be vanishing across the pond as well. In a 6.6-million-word corpus of British English (MacWhinney 2000), not a single instance of *shan't* is found. And its frequency of usage has been in a steady decline since 1800, the beginning date of the Google Books Corpus. As of *daren't*, the OED does not provide any citation and it has always been very rare throughout the period of the Google Books Corpus. These gapped forms are marked by ∅.

The prescriptively maligned *ain't* ([eɪnt]), however, is robustly attested for *am*, *are*, *is*, *have*, and *has* in COCA as well as a six-million-word corpus of child-directed American English (MacWhinney 2000). Since the phonological form of [eɪnt] is unpredictable from the auxiliary host, it is boldfaced in Table 2 to mark its idiosyncrasy, along with a few other exceptions to which I return later. Note that the frequency estimates of the *ain't* forms are approximate. First, I only counted strings where *ain't* is immediately preceded by a pronoun – the majority case, but sentences with a lexical subject (e.g., "Kids ain't ready") are not included. Second, because both *be* and *have* can take on *ain't*, the counts

Table 2: The morphophonological alternation of n’t contraction

| aux+not | | n’t contraction | | (%) |
|---|---|---|---|---|
| could [kud] | 45,256 | [kudn̩t] | 106,123 | 70.104 |
| did [dɪd] | 128,432 | [dɪdn̩t] | 342,202 | 72.711 |
| does [dʌz] | 72,194 | [dʌzn̩t] | 164,922 | 69.553 |
| had [hæd] | 27,410 | [hædn̩t] | 46,987 | 63.157 |
| has [hæz] | 28,529 | [hæzn̩t] | 29,578 | 50.255 |
| has [hæz] | 28,529 | **[eɪnt]** | 749 | 1.273 |
| have [hæv] | 24,957 | [hævn̩t] | 45,849 | 63.868 |
| have [hæv] | 24,957 | **[eɪnt]** | 981 | 1.367 |
| is [ɪz] | 189,538 | [izn̩t] | 100,164 | 34.275 |
| is [ɪz] | 189,538 | **[eɪnt]** | 2,537 | 0.868 |
| might [maɪt] | 14,780 | [maɪtn̩t] | 78 | 0.525 |
| must [mʌst] | 4,156 | **[mʌsn̩t]** | 917 | 18.076 |
| need [nid] | 3,705 | [nidn̩t] | 1,235 | 25.000 |
| ought [ɔt] | 1,031 | [ɔtn̩t] | 66 | 6.016 |
| should [ʃud] | 20,577 | [ʃudn̩t] | 25,576 | 55.416 |
| was [wʌz] | 97,457 | [wʌzn̩t] | 141,384 | 59.196 |
| would [wud] | 46,205 | [wudn̩t] | 85,853 | 65.012 |
| am [æm] | 10,258 | ∅ | 5 | 0.041 |
| am [æm] | 10,258 | **[eɪnt]** | 2,046 | 16.622 |
| are [ar] | 89,083 | [arnt] | 50,137 | 35.602 |
| are [ar] | 89,083 | **[eɪnt]** | 1,073 | 0.765 |
| can [kæn] | 75,531 | **[kænt]** | 201,060 | 72.692 |
| dare [dɛər] | 320 | ∅ | 25 | 7.246 |
| do [du] | 81,074 | **[dont]** | 654,576 | 88.979 |
| may [meɪ] | 36,195 | ∅ | 12 | 0.033 |
| shall [ʃæl] | 1,271 | ∅ | 123 | 8.824 |
| were [wr̩] | 41,224 | [wr̩nt] | 35,120 | 46.002 |
| will [wɪl] | 39,068 | **[wont]** | 86,158 | 68.802 |

for the auxiliaries are parceled out by extrapolating from the frequencies of the regularly contracted *n’t* forms.[2] For instance, there are 2,054 instances of “you/they ain’t”: the “share” for *are* is based on the count of “aren’t” (50,137) relative to “haven’t” (45,849).

[2] Here I gloss over the fact that there are English dialects in which *ain’t* is also an alternative form of negative contraction for *do*, *does*, and *did* (e.g., Labov et al. 1968; Weldon 1994). It would be difficult to estimate their frequencies but formally, this use of *ain’t* serves to create additional (unpredictable) exceptions to the contraction process which, as we discuss below, contributes to the breakdown of productivity and the emergence of gaps.

This amounts to 52.2% of 2,054, or 1,073, as recorded in the Table. Finally, the estimate of *ain't* as the contraction of *am* + *n't* cannot follow a similar process because, of course, *amn't* is gapped. I thus allocated roughly 75% of the "I ain't" counts, which is the share of "I am not" out of the total of "I am not" and "I have not", to the contraction of *am not*. For these five auxiliaries that can be realized as *ain't*, the percentage of the contracted forms are based on the sum of uncontracted, *n't*-contracted, and *ain't*-contracted forms. More precise estimates are certainly possible but as we will see, the exact frequencies are not especially important for our purposes: it is more pertinent to approximate a "typical" English speaker's experience with these forms. Roughly, we would like to know whether an English speaker will have encountered a specific phonological word at all, by using some independently motivated frequency threshold (e.g., once per million; Nagy & Anderson 1984): it is evident that the frequency of *ain't* is sufficiently high for this threshold despite our rough estimates.

A tempting approach to gaps is to appeal to indirect negative evidence (Chomsky 1981; Pinker 1989). A strong version takes the shape of *lexical conservatism*: do not use a form unless it is explicitly attested. This recalls Halle's [-Lexical Insertion] treatment of gaps in his classic paper (1973) and can be found in recent works as well (e.g., Pertsova 2005; Steriade 1997; Rice 2005; Wolf & McCarthy 2009). A weak version makes use of frequency information. For instance, if *amn't* were possible, language learners would have surely heard it in the input, especially since *am* is highly frequent and would have had plenty of opportunities to undergo *n't* contraction. Its conspicuous absence, then, would provide evidence for its ungrammaticality (e.g., Daland, Sims & Pierrehumbert 2007; Sims 2006; Baerman 2008; Albright 2009).

Traditional acquisition research has always viewed indirect negative evidence with strong suspicion (Berwick 1985; Osherson, Stob & Weinstein 1986; Pinker 1989). Research on the *amn't* gap (e.g. Hudson 2000) has also questioned its usefulness. However, with the recent rise of probabilistic approaches to language acquisition especially Bayesian models of inference, the field has seen a revival of indirect negative evidence. If the conception of learning is a zero-sum – or more precisely, one-sum — game which assigns a probabilistic distribution over all linguistic forms, the unattested will necessarily lose out to the attested, at least in most probabilistic models of language learning. A thorough assessment of indirect negative evidence within a probabilistic framework is beyond the scope of the present paper; see Niyogi 2006; Yang 2015; Yang et al. 2017. But a careful statistical examination of gaps serves to reveal its deficiencies. Note that the question is not whether indirect negative evidence can account for *some* missing forms: the absence of *amn't* is indeed unexpected under any reasonable formulation. The real challenge is to ensure that indirect negative evidence will pick out *only* the gapped forms but nothing else, while keeping in mind that morphological inflection is generally not gapped but fully productive, readily extending to novel items.

Two observations can be made about the frequency statistics in Table 2, which suggest that indirect negative evidence is unlikely to succeed. First the *n't* forms of several auxiliaries such as *might* and *need* are in fact quite rare. They appear considerably less frequently than once per million, which is generally regarded as the minimum threshold to guarantee exposure for most English speakers (Nagy & Anderson 1984). In the six-million-word corpus of child-directed American English (MacWhinney 2000), *mightn't* appears only once, *needn't* appears only twice, and *mustn't* does not appear at all. In the other words, these *n't* forms may be so rare that they are in effect absent for many children (Hart & Risley 1995). Lexical conservatism thus will not distinguish them from the truly gapped *amn't*, *mayn't*, *daren't*, and *shan't*, the last of which is in fact more frequently attested in COCA. Second, consider a statistical interpretation of indirect negative evidence. The last column of Table 2 provides the percentage of the *n't* contraction out of all negated forms. An auxiliary with an unusually low ratio may mean that it has performed below expectation and could be a clue for its defectiveness. However, the statistics in Table 1 suggest otherwise. It is true that *amn't* and *mayn't* have very low ratios: this fact alone is not remarkable because these are indeed gaps. But exactly how low should a ratio be for the learner to regard a contracted form to be defective? On the one hand, we have *mightn't* and *oughtn't* at 0.525% and 6.016%, and these are not defective. On the other hand, we have *daren't* and *shan't* at 7.246% and 8.824%, but these in fact are defective. There doesn't appear to be a threshold of frequency or probability that can unambiguously distinguish gapped from ungapped items.

## 3.3 N't Contraction in Language Development and Change

Let's see how the Tolerance Principle provides an account of the *amn't* gaps. The simplest approach is to consider all the auxiliary verbs and their *n't* contractions collectively as a homogeneous set. Using the once-per-million threshold as a reasonable approximation of a typical American English speaker's vocabulary, and taking the size of the Corpus of Contemporary American English (520 million words) into account, there are 18 auxiliaries with reliably attested *n't* forms. The four gapped forms are all below this threshold and are thus excluded from consideration. It is important to clarify that, unlike various forms of lexical conservatism and indirect negative evidence discussed earlier, we do not regard the absence of these forms as evidence for their defectiveness. Rather, the learner's task is to deduce, on the basis of the 18 well-attested forms, including *am~ain't*, that *n't* contraction is not a productive pattern in the English auxiliary system.

This is quite easily accomplished. Of the 18 auxiliaries, *n't* is realized as follows:

(5)
   a. [n̩t]: could, did, does, had, need, should, was, would (8)
   b. [eɪnt] in variation with either [nt] or [n̩t]: have, has, is, are (4)
   c. [nt]: can, were (2)
   d. idiosyncratic vowel change: do, will (2)
   e. [eɪnt]: am (1)
   f. [n̩t] but idiosyncratically deletes [t] in the auxiliary (see Zwicky & Pullum 1983: 508–509 for discussion): must (1)

For any of these alternations to be productive, it must have no more than $\theta_{18} = 6$ exceptions. The most promising [n̩t], which applies to 8 auxiliaries and thus has 10 exceptions, is a long way off. Even if we are to include the [n̩t]-taking *have*, *has*, and *is* and ignore the unpredictable variant [eɪnt] form, the rule "nt ⟶ n̩t" still falls short of productivity. Thus, the learner will be able to conclude, from the Conditions on Gaps (4), that *n't* contraction is not a productive process for English auxiliaries and must be learned lexically. If *amn't* fails to appear in the input, it will be absent. Only after the learner has already concluded that a category does not have a productive rule can they start to regard the absence of evidence as evidence of absence.

The preceding analysis, while correctly identifies the *n't* gaps, has some inadequacies. For one thing, based on the 18 contracted forms, the primary evidence for language acquisition, learners would also identify *mustn't* and *oughtn't* as gapped as they fall below the minimum frequency of once per million. This is not necessarily a fatal shortcoming: *mustn't* and *oughtn't* are still considerably more frequent than *amn't* and *mayn't*, the two genuinely gapped forms, and children may acquire them in later stages of acquisition. But more significantly, as Steve pointed out to me in a personal communication (unrelated to the current celebratory volume), the preceding brute-force approach misses an important structural generalization. Table 2 is divided into two halves on Steve's advice. As he insightfully observes, none of the auxiliaries that ends in an obstruent is gapped; these are listed in the top portion of the Table. By contrast, gaps are only found in the auxiliaries that do not end in an obstruent, which are listed in the bottom portion of the Table.

If we carry out a Tolerance analysis along the feature [±obstruent], a much more elegant and interesting pattern emerges. For the 12 [+obstruent] auxiliaries, only four have exceptions – *has*, *have*, *is*, and *must* – just below $\theta_{12} = 4$. Thus, English learners can identify a productive rule:

(6) nt ⟶ n̩t / [+obstruent] # ___

This immediately accounts for the fact that speakers generally accept the forms *mightn't* and *oughtn't* despite their very low frequencies (well below once per million): these two auxiliaries, of course, follow the structural description of (6). By contrast, *amn't*, *mayn't*, *daren't*, and *shan't*, some of which appear more frequently than *mightn't* and *oughtn't*, are generally rejected because they fail to meet the structural descriptions of the productive rule in (6).

Consider now the six [-obstruent] auxiliaries in the bottom portion of Table 2. Here *am* and *are* have [eɪnt], *can* and *were* add [nt], and *do* and *will* have idiosyncratic vowel changes. Since the Tolerance threshold $\theta_6 = 3$, no distinct pattern will be identified as productive: lexicalization is required and gaps are predicted for *mayn't*, *daren't*, *shalln't*, and of course *amn't*.

The calculation here is very delicate but it is interesting to push the Tolerance Principle to the limit. What if the child has not learned *ain't* as the *n't*-contracted form for *am* and *are*? Although *ain't* forms are quite robustly attested in COCA as well as in child-directed English, they are still strongly dialectal and are, at least in the input to some children, less frequent than the "regular" forms such as *aren't*, *isn't*, *haven't*, and *hasn't*. If so, a child during an early stage of acquisition may in effect have only five [-obstruent] auxiliaries and their contracted forms to learn from: namely, *are*, *can*, *do*, *were*, and *will*. Here the statistically dominant pattern of "nt ⟶ [nt] / [-obstruent] # ___" does reach productivity: the two idiosyncratic exceptions of *do* and *will* fall below the threshold of $\theta_5 = 3$, and *n't* contraction is predicted to be transiently productive!

Bill Labov (personal communication) distinctly recalls being a young *amn't* speaker only to exit that stage at a later time. Indeed, we can find attested examples in American English-learning children's speech. The three examples in (7) are taken from the CHILDES database (MacWhinney 2000):

(7) a. I amn't a dad. (Kate/Kim, 3;6: Sawyer Corpus 3-12-92.cha)
b. I'm doing this puzzle well, amn't I? (Mark, 3;11: MacWhinney Corpus 67b1.cha)
c. Amn't I clever? (Mark, 3;11: MacWhinney Corpus 67b2.cha)

The reader is encouraged to listen the audio recordings of the examples in (7) in the CHILDES database. The first child's identity is unclear due to discrepancies in transcription. The examples from Mark can be heard as the investigator's exact revoicing (Brian MacWhinney, personal communication). Although three examples seem quite rare, it is worth noting that almost all *am*'s are contracted onto the pronoun (i.e., *I'm not*). Of the one million American English child utterances, there are only 42 full forms of *am* followed by negation (i.e., *I am not*), which makes the three *amn't* errors not so negligible. Of course, everyone eventually hears 'I ain't': from pop songs on radio if not from the immediate family and friends. Thus, *amn't* will disappear according to the Tolerance-based analysis, for *ain't* introduces an additional exception which leads to the breakdown of productivity for the [-obstruent] class.

Corroborative evidence for the (transient) productivity of *n't* contraction can also be found in other auxiliaries. To my great surprise, there are numerous instances of *willn't* as the negative contracted form of *will* and *whyn't* for 'why don't/didn't' in the speech of many parent-child dyads, apparently all from the New England region. Other than enriching the empirical data on contraction, *willn't* and *whyn't* do not tell us much about the productivity of *n't* contraction or its acquisition: if parents use them frequently, and they do, children will follow. Nevertheless, *willn't* can also be found in the spontaneous speech of children who are not from the New England region:[3]

(8) a. No we willn't. (Ross 2;9, Colorado, MacWhinney Corpus 26b2.cha)
b. Oh it willn't fit in there (Marie 6;6, Ontario, Evans Corpus dyad07.cha)
c. He willn't be a good boy (Jared 6;7, Ontario, Evans Corpus dyad19.cha)

Perhaps most strikingly is an utterance produced by Sarah, a child from the Harvard studies (Brown 1973):[4]

(9) And the reindeer saidn't.

[3] Brain MacWhinney (personal communication) confirmed that the only time he or his wife ever used *willn't* was when transcribing Ross's speech.

[4] The contraction of *n't* onto the main verb as in (9) was attested in the history of English: see Brainerd 1989 for *caren't* ('don't care') and Jespersen 1917 for *bettern't*, *usen't*, and indeed *whyn't*.

Taken together, the examples in (7), (8), and (9) suggest that *n't* contraction is at least transiently productive for some English-learning children.

Ross's *willn't* presents an especially interesting opportunity for studying the productivity of *n't* contraction. The CHILDES corpus contains a relatively extensive record of Ross's longitudinal language development. We can then study his auxiliaries and contractions, and subject his individual grammar to the kind of fine-grained analysis of Adam's past tense (§2). By the time Ross produced *No we willn't*, he had used 9 *n't*-contracted auxiliaries:

(10) a. couldn't, didn't, haven't, isn't, wouldn't
 b. aren't, can't, don't, won't

If Ross had not started partitioning the auxiliaries by the [±obstruent] feature, the $N = 9$ examples in (10) supports the productive use of *n't* contraction because the four examples in (10b) are below the number of tolerable exceptions ($\theta_9 = 4.2$). The 5/4 split between rule-governed and exceptional items is exactly the stimuli used in the artificial language study (Schuler, Yang & Newport 2016) where children nearly categorically generalized the rule. If he failed to distinguish the syllabic [n̩] in (10a) and the nonsyllabic [n] in *aren't* and *can't* in (10b), it would have been even easier for *n't* contraction to reach productivity. Thus, Ross's productive use of *n't* contraction in (8) is predicted by the Tolerance Principle.

The naturalistic evidence from child language is admittedly thin, but it suggests that the emergence of the *amn't* and other gaps in the auxiliary system may be due to the use of *ain't*. Again, the gaps would not be the result of mutual exclusivity: there are doublets such as *haven't*~*ain't* etc. so *amn't* and *ain't* could have coexisted side by side. Gaps arise/arose because the form of *ain't* weakens the numerical advantage of *n't* contraction, pushing it below the Tolerance threshold.

Finally, a little historical detective work bolsters our treatment of the *amn't* gap.[5] According to Jespersen (1917: 117), "the contracted forms seem to have come into use in speech, though not yet in writing, about the year 1600." The change appears to have originated in non-standard speech before spreading to mainstream usage. Subsequent scholarship, however, places the date to a somewhat later time (e.g., 1630, Brainerd 1989: 181; see also Warner 1993: 208–209). Pursuing the results from the Tolerance-based analysis, we can make two observations.

First, it is likely that *n't*-contraction was at one point productive, which seems especially effective for the [+obstruent] auxiliaries; see also (9) and fn 4. Brainerd's study finds that *didn't*, *hadn't*, *shouldn't*, and *wouldn't* appeared from 1670s, soon after the *n't* contraction appeared in the English language. These were followed by *couldn't*, *mightn't*, *needn't*, and *mustn't* in the 18th century, and the last to join the group was *oughtn't* in the 19th century, first attested in Dicken's 1836 *The Village Coquette*. Thus speakers at that time must have formed a productive contraction rule for [+obstruent] auxiliaries, perhaps like the one given in (6). Following this line of reasoning, we make the pre-

[5] I am grateful to Anthony Warner for pointing out the important study of Brainerd 1989.

diction, admittedly one that is difficult to test, that if a new [+obstruent] auxiliary is to appear in the language, it will be immediately eligible for *n't* contraction.

Second, and in contrast to the [+obstruent] class that had been expanding the number of *n't* contractible auxiliaries, the [-obstruent] class has been steadily losing members. Interestingly, the [-obstruent] auxiliaries were quite systematically available for *n't* contraction by the end of the 17th century (Brainerd 1989). Of special interest are of course those that were *n't* contracted in the past but are presently gapped. According to Brainerd's study, the first instance of *shan't* appeared in 1631, *mayn't* in 1674, *daren't* in 1701: all three are now gapped. The very fact that they fall out of usage points to the non-productivity of *n't* contraction for these [-obstruent] auxiliaries: in general, a productive rule would have gained rather than lost members.

How, we wonder, did the [-obstruent] class lose its productivity? Much more detailed historical investigation will be needed but an interesting hypothesis can be offered as follows. The historical development of *n't* contraction may mirror the trajectory of language acquisition by children; that is, ontogeny may recapitulate phylogeny. Our discussion of children's *n't* contraction in modern American English suggests that the use of *ain't* for *am not*, which children probably acquire later during acquisition, increases the number of exceptions for the contraction process. It is conceivable that the emergence of *ain't*, an unpredictably contracted form of *am not*, was also the culprit for the breakdown of productivity.

Historically, *an't*/*a'nt* surfaced as the contracted form of *am not* between 1673 and 1690. But by the early 1700s, *an't*/*a'nt* began to be used for both *am not* and *are not* (Brainerd 1989: 186). Whatever the phonological cause for this convergence, or how/when *ain't* joined the fray, the effect is that *am not* no longer had a predictable form of contraction. If our analysis of children's *amn't* and *willn't* is correct, then we would find *amn't* and *ain't* to be in complementary distribution: If a dialect does not allow *ain't* for *am not*, *amn't* would be possible; otherwise *amn't* would be gapped.

The most direct evidence for this suggestion comes from the dialectal distribution of *amn't*, and its correlation with *ain't*. The OED notes that *amn't* is present in "nonstandard" American English and various northern parts of the UK. There is little to suggest that *amn't* is possible in American English at all; all the five occurrences in COCA come from Scottish and Irish writers.[6] It is remarkable, then, that Scotland and Ireland have "traditionally completely *ain't*-free dialects" (Anderwald 2003b: 520): it is precisely in these regions where *amn't* is robustly attested, both in the century-old *The English Dialect Dictionary* (Wright 1898) and in recent dialect surveys of English (Anderwald 2003a).[7]

Before I conclude this section, it is important to clarify the scope of the present analysis. The Tolerance Principle, through Conditions on Gaps (4), can identify defective morphological category where gaps *may* emerge. Such categories are defined by the

[6] The corpus of child-directed American-English, surprisingly, contains one instance of *amn't*: "I am stirring, amn't I?" It was produced by Colin Fraser, on staff in Roger Brown's Harvard study of language acquisition (1973). A little Internet research reveals that Fraser, later a Cambridge scholar with a few psychology textbooks to his credit, is a native of Aberdeen.

[7] I would like to thank Gary Thoms for discussion for the distribution of *amn't* in Scottish English.

structural descriptions of rules. It does not predict, at least synchronically, which items within these categories will be gapped. That issue, in my view, is completely a matter of usage frequency: if the inflected form of an item in a defective category is used very rarely or not at all, it will be gapped. Of course, it is also possible that no gaps are found in a defective morphological category, if all items happen to be inflected sufficiently frequently. In that case, however, we do predict that if a novel item matches the structural description of a defective category, the speaker will be at a loss to produce an inflected form. Thus, the emergence of gaps, just as the calibration of productivity, is determined by the composition of the input data. Finally, the preliminary work on the history of *n't* contraction suggests that the Tolerance Principle can be applied to the study of language change. It makes concrete predictions about productivity – the rules that could gain new members, and the rules that could only lose existing members – as long as the relevant values of $N$ and $e$ from historical data can be reliably estimated. The reader is referred to Yang 2016 for a case study of the so-called dative sickness in Icelandic morphosyntax.

## 4 Gaps in I-language

Halle's classic paper (1973) contains the much criticized proposal that gaps are caused by the [+Lexical Insertion] feature associated with certain forms. As noted earlier, this kind of lexical conservatism is difficult to reconcile with the unbounded generativity of word formation, and similar approaches using indirect negative evidence are also unlikely to succeed. But in a footnote of that very paper, Halle proposes an alternative approach which he himself regards as equivalent but has almost never been discussed by other researchers:

> The proposal just sketched might be modified somewhat as regards the treatment of words formed by rules that traditionally have been called "nonproductive". One might propose that all words formed by non-productive rules are marked by these rules as [-Lexical Insertion]. The smaller subset of actually occurring words formed by such rules would then be listed in the filter with the feature [+Lexical Insertion]. … In other words, it is assumed that words generated by a productive process are all actually occurring and that only exceptionally may a word of this type be ruled out of the language. On the other hand, words generated by a nonproductive rule are assumed not to be occurring except under special circumstances. In this fashion we might capture the difference between productive and nonproductive formations (5).

Hetzron (1975), while arguing against Halle's [+Lexical Insertion] proposal, makes essentially the same suggestion. Rules are either productive or lexicalized, and gaps arise in the unproductive corners of the grammar. His conception of gaps can be strongly identified with the Elsewhere Condition, a critical component of the present theory:

> The speaker must use ready-made material only for "exceptional" forms, while everywhere else he could very well "invoke the word formation component". Technically, this can be represented by a disjunctive set of rules where idiosyncratic or

> "exceptional" formations are listed with as much explicitness as necessary, while the general word formation rules would appear afterward, with the power to apply "to the rest" (871).

That is, gaps arise when productivity fails. The problem of gaps thus reduces to the problem of productivity. Some subsequent proposals have adopted a similar approach (Albright 2009; Baronian 2005; Hudson 2000; Maiden & O'Neill 2010; Pullum & Wilson 1977; Sims 2006), including Steve's own account (2010): gaps result from conflicting forces in word formation such that the output form becomes unpredictable and thus unrealized. The Tolerance Principle provides a precise solution of what makes a rule productive, and its application to gaps reinforces the general position that gaps and productivity are two sides of the same coin.

The Tolerance Principle is a provable consequence of the Elsewhere Condition and follows from the general principle of efficient computation: the child prefers *faster* grammars, a "third factor" in language design par excellence (Chomsky 2005). In fact, a stronger claim can be made in favor of such an analytical approach. I submit that a descriptive analysis of languages, however typologically complete or methodologically sophisticated, cannot in principle provide the right solution for productivity. First, as noted earlier, the categorical nature of children's morphological acquisition suggests that productivity must be demarcated by a discrete threshold (see also Aronoff 1976: 36). But note that such a threshold is empirically *undiscoverable*. Productive processes will lie above the threshold and unproductive processes will lie below, but with arbitrary "distance" from it in both cases. Thus, the threshold cannot be regressed out of the data. Second, while linguists now have an ever expanding arsenal of investigative tools to study productivity, ranging from the Wug test to fMRI to Big Data, the psychological grammar is developed without supervision in a matter of few years; these new empirical methods presently are at best a description of the speaker's grammatical knowledge and not yet learning models that account for how such knowledge is acquired. Finally, even if we were to discover the threshold of productivity through a statistical analysis – e.g., a productive rule must hold for at least 85% of eligible words – it would still remain mysterious why the critical value is exactly what it is, rather than 80% or 90%.

In other words, an I-language approach to productivity is needed, one which builds exclusively on the inherent constraints on language and cognition that all children have access to, with deductively established properties that must hold universally across languages. The study of language as a part of human biology, I believe, is an approach that Steve endorses and pursues (Anderson & Lightfoot 2002), which can be seen in his writings on morphology and related issues (Anderson 2010a; 2015).

Finally, a personal note. It is no exaggeration to say that I owe my professional career to Steve. He managed to create a position for me at Yale, which kept me close to my young family and thus linguistics, and further away from the seductive fortunes in the tech sector. It was also Steve who taught me, more effectively than anyone, the difference between linguistic evidence and rhetoric. It has been a privilege to learn from him. To figure out how to wake up irregular took over 15 years; the answer, I hope, is to his satisfaction. It may once again win me a spot, this time in the Linguistic Club of Ashville, North Carolina.

## Acknowledgements

I would like to thank Steve Anderson, Bill Labov, Brian MacWhinney, Gary Thoms, Anthony Warner, and an anonymous reviewer for helpful discussions and comments.

## References

Albright, Adam. 2009. Lexical and morphological conditioning of paradigm gaps. In Curt Rice & Sylvia Blaho (eds.), *Modeling ungrammaticality in Optimality Theory*, 117–164. London: Equinox.

Anderson, Stephen R. 1969. *West Scandinavian vowel systems and the ordering of phonological rules*. Cambridge, MA: Massachusetts Institute of Technology PhD thesis.

Anderson, Stephen R. 1973. Remarks on the phonology of English inflection. *Language and Literature* 1(4). 33–52.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. 2008. Phonologically conditioned allomorphy in the morphology of Surmiran (Rumantsch). *Word Structure* 1(2). 109–134.

Anderson, Stephen R. 2010a. An I-language view of morphological 'exceptionality': Comments on Corbett's paper. In Horst J. Simon & Heike Wiese (eds.), *Expecting the unexpected: Exceptions in grammar*, 127–134. Berlin: Mouton de Gruyter.

Anderson, Stephen R. 2010b. Failing one's obligations: Defectiveness in Rumantsch reflexes of DĒBĒRE. In Matthew Baerman, Greville G. Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing forms and what they tell us*, 19–34. Oxford: British Academy & Oxford University Press.

Anderson, Stephen R. 2015. Morphological change. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 264–285. New York: Routledge.

Anderson, Stephen R. & David W. Lightfoot. 2002. *The language organ: Linguistics as cognitive physiology*. Cambridge: Cambridge University Press.

Anderwald, Lieselotte. 2003a. *Negation in non-standard British English: Gaps, regularizations and asymmetries*. London: Routledge.

Anderwald, Lieselotte. 2003b. Non-standard English and typological principles: The case of negation. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 507–530. Berlin/New York: Mouton de Gruyter.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Baerman, Matthew. 2008. Historical observations on defectiveness: The first singular non-past. *Russian Linguistics* 32(1). 81–97.

Baerman, Matthew, Greville G. Corbett & Dunstan Brown (eds.). 2010. *Defective paradigms: Missing forms and what they tell us*. Oxford: British Academy & Oxford University Press.

Baker, Robert & Philip Smith. 1976. A psycholinguistic study of English stress assignment rules. *Language and Speech* 19(1). 9–27.

Baronian, Luc. 2005. *North of phonology*. Stanford University PhD thesis.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14(2–3). 150–177.
Berwick, Robert. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.
Berwick, Robert C & Noam Chomsky. 2016. *Why only us: Language and evolution*. Cambridge, MA: MIT Press.
Bloch, Bernard. 1947. English verb inflection. *Language* 23(4). 399–418.
Bowerman, Melissa. 1982. Reorganizational process in lexical and syntactic development. In Eric Wanner & Lila R. Gleitman (eds.), *Language acquisition: The state of the art*, 319–346. Cambridge/New York: Cambridge University Press.
Brainerd, Barron. 1989. The contractions of not: A historical note. *Journal of English Linguistics* 22(2). 176–196.
Bresnan, Joan. 2001. Explaining morphosyntactic competition. In Mark Baltin & Chris Collins (eds.), *Handbook of contemporary syntactic theory*, 1–44. Malden, MA & Oxford: Blackwell Publishing.
Broadbent, Judith M. 2009. The **amn't* gap: The view from West Yorkshire. *Journal of Linguistics* 45(2). 251–284.
Brown, Dunstan & Andrew Hippisley. 2012. *Network morphology: A defaults-based theory of word structure*. Vol. 133. Cambridge: Cambridge University Press.
Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.
Brown, Roger, Courtney B. Cazden & Ursula Bellugi. 1973. The child's grammar from I to III. In Charles A. Ferguson & Daniel I. Slobin (eds.), *Studies of child language development*, 295–333. New York: Holt, Rinehart & Winston.
Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Philadelphia: John Benjamins.
Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms., Harvard University and MIT. Revised version published by Plenum, New York, 1975.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1981. *Lectures in government and binding*. Dordrecht: Foris.
Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36. 1–22.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. Cambridge, MA: MIT Press.
Clahsen, Harald. 1999. Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences* 22. 991–1069.
Clahsen, Harald, Monika Rothweiler, Andreas Woest & Gary Marcus. 1992. Regular and irregular inflection in the acquisition of German noun plurals. *Cognition* 45. 225–255.
Cutler, Anne & David M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2(3–4). 133–142.
Daland, Robert T., Andrea D. Sims & Janet Pierrehumbert. 2007. Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Proceedings of the 45th annual meeting of the Association for Computational Linguistics*, 936–943. Somerset, NJ: Association for Computational Linguistics.
Davies, Mark. 2008. Corpus of Contemporary American English (COCA): 410+ million words, 1990-present. Available at http://www.americancorpus.org/.

Frampton, John. 2001. The *amn't* gap, ineffability, and anomalous *aren't*: Against morphosyntactic competition. In Mary Andronis, Christopher Ball, Heidi Elston & Sylvain Neuvel (eds.), *Papers from the 37th meeting of the Chicago Linguistic Society*, 399–412. Chicago: Chicago Linguistic Society.

Guion, Susan, J.J. Clark, Tetsuo Harada & Ratree Wayland. 2003. Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46(4). 403–427.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4(1). 3–16.

Halle, Morris. 1998. The stress of English words 1968-1998. *Linguistic Inquiry* 29(4). 539–568.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress.* Cambridge, MA: MIT Press.

Hart, Betty & Todd R Risley. 1995. *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes Publishing.

Hay, Jennifer & R. Harald Baayen. 2003. Phonotactics, parsing, and productivity. *Italian Journal of Linguistics* 1. 99–130.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13(2). 227–276.

Hetzron, Robert. 1975. Where the grammar fails. *Language* 51(4). 859–872.

Hudson, Richard. 2000. *I amn't. *Language* 76(2). 297–323.

Jespersen, Otto. 1917. *Negation in English and other languages* (Historisk-flologiske Meddelelser 1). Copenhagen: A.F. Høst.

Kehoe, Margaret & Carol Stoel-Gammon. 1997. The acquisition of prosodic structure: An investigation of current accounts of children's prosodic development. *Language* 73. 113–144.

Kelly, Michael H. 1992. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review* 99(2). 349–364.

Kiparsky, Paul. 1973. Elsewhere in phonology. In Stephen R. Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 93–106. New York: Holt, Rinehart & Winston.

Labov, William, Paul Cohen, Clarence Robins & John Lewis. 1968. A study of the nonstandard English of Negro and Puerto Rican speakers in New York City. Volume 1: Phonological and grammatical analysis. Cooperative Research Project No. 3288, Columbia University, New York.

Ladefoged, Peter & Victoria Fromkin. 1968. Experiments on competence and performance. *IEEE Transactions on Audio and Electroacoustics* 16(1). 130–136.

Legate, Julie A. & Charles Yang. 2013. Assessing child and adult grammar. In Robert Berwick & Massimo Piattelli-Palmarini (eds.), *Rich languages from poor inputs: In honor of Carol Chomsky*, 168–182. Oxford: Oxford University Press.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk.* 3rd. Mahwah, NJ: Lawrence Erlbaum.

Maiden, Martin & Paul O'Neill. 2010. On morphomic defectiveness: Evidence from the Romance languages of the Iberian Peninsula. In Matthew Baerman, Greville G. Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing forms and what they tell us*, 103–124. Oxford: British Academy & Oxford University Press.

Marcus, Gary, Steven Pinker, Michael T. Ullman, Michelle Hollander, John Rosen & Fei Xu. 1992. *Overregularization in language acquisition*. Vol. 57 (Monographs of the Society for Research in Child Development 4). Chicago: University of Chicago Press.

McClelland, James L. & Karalyn Patterson. 2002. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences* 6(11). 465–472.

Nagy, William E. & Richard C. Anderson. 1984. How many words are there in printed school English? *Reading Research Quarterly* 19(3). 304–330.

Newport, Elissa. 1990. Maturational constraints on language learning. *Cognitive Science* 14(1). 11–28.

Nida, Eugene A. 1949. *Morphology: The descriptive analysis of words*. 2nd. Ann Arbor: University of Michigan Press.

Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

Osherson, Daniel N., Michael Stob & Scott Weinstein. 1986. *Systems that learn: An introduction to learning theory for cognitive and computer scientists*. Cambridge, MA: MIT Press.

Pertsova, Katya. 2005. How lexical conservatism can lead to paradigm gaps. In Jeffrey Heinz, Andrew Martin & Katya Pertsova (eds.), *UCLA working papers in linguistics 11: Papers in phonology 6*, 13–30. Los Angeles: UCLA Linguistics Department.

Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

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

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Cambridge, MA: MIT Press.

Pullum, Geoffrey K. & Deidre Wilson. 1977. Autonomous syntax and the analysis of auxiliaries. *Language* 53(4). 741–788.

Rice, Curt. 2005. Optimal gaps in optimal paradigms. *Catalan Journal of Linguistics* 4. 155–170.

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

Schuler, Kathryn, Charles Yang & Elissa Newport. 2016. Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting*. Philadelphia, PA.

Schütze, Carson T. 2005. Thinking about what we are asking speakers to do. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 457–485. Berlin: Mouton de Gruyter.

Sims, Andrea D. 2006. *Minding the gap: Inflectional defectiveness in a paradigmatic theory*. Ohio State University PhD thesis.

Steriade, Donca. 1997. Lexical conservatism. In Linguistic Society of Korea (ed.), *Linguistics in the morning calm 4: Selected papers from SICOL–1997*, 157–179. Seoul: Hanshin.

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

Warner, Anthony. 1993. *English auxiliaries: Structure and history*. Cambridge University Press.

Weldon, Tracey. 1994. Variability in negation in African American Vernacular English. *Language Variation and Change* 6(3). 359–397.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Wolf, Matthew & John J. McCarthy. 2009. Less than zero: Correspondence and the null output. In Curt Rice & Sylvia Blaho (eds.), *Modeling ungrammaticality in Optimality Theory*, 17–66. London: Equinox.

Wright, Joseph. 1898. *The English dialect dictionary*. London: Henry Frowde.

Xu, Fei & Steven Pinker. 1995. Weird past tense forms. *Journal of Child Language* 22(3). 531–556.

Yang, Charles. 2002a. A principle of word storage. Manuscript: Yale University.

Yang, Charles. 2002b. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang, Charles. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8(10). 451–456.

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

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

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

Yang, Charles, Stephen Crain, Robert C. Berwick, Noam Chomsky & Johan J. Bolhuis. 2017. The growth of language: Universal Grammar, experience, and principles of efficient computation. *Neuroscience and Biobehavioral Review*.

Yip, Kenneth & Gerald J. Sussman. 1998. *Sparse representations for fast, one-shot learning*. Tech. rep. 1633. MIT Artificial Intelligence Laboratory.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English N'T. *Language* 59(3). 502–513.